Does DNA act as a telephone line? *p. 1284*

Mutations enhancing leukemia development *pp. 1291 & 1373*

A stable gold support gives sharper resolution *p. 1377*

Science

$10
12 DECEMBER 2014
sciencemag.org

AAAS



SPECIAL ISSUE

Avian genomes

Sequencing across the bird species tree

p. 1308

Science
AAAS
1200 New York Avenue, NW, Washington, DC 20005
Bateman House, 82-88 Hills Road, Cambridge, UK CB2 1LQ

Editor-in-Chief Marcia McNutt
Executive Editor Monica M. Bradford **News Editor** Tim Appenzeller
Managing Editor, Research Journals Katrina L. Kelner
Deputy Editors Barbara R. Jasny, Andrew M. Sugden(UK), Valda J. Vinson, Jake S. Yeston

Research and Insights

SR. EDITORS Caroline Ash(UK), Gilbert J. Chin, Lisa D. Chong, Maria Cruz(UK), Julia Fahrenkamp-Uppenbrink(UK), Pamela J. Hines, Stella M. Hurtley(UK), Paula A. Kiberstis, Marc S. Lavine(Canada), Kristen L. Mueller, Ian S. Osborne(UK), Beverly A. Purnell, L. Bryan Ray, Guy Riddihough, H. Jesse Smith, Jelena Stajic, Peter Stern(UK), Phillip D. Szuromi, Brad Wible, Nicholas S. Wigginton, Laura M. Zahn **ASSOCIATE EDITORS** Brent Grocholski, Melissa R. McCartney, Margaret M. Moerchen, Sacha Vignieri **ASSOCIATE BOOK REVIEW EDITOR** Valerie B. Thompson **ASSOCIATE LETTERS EDITOR** Jennifer Sills **CHIEF CONTENT PRODUCTION EDITOR** Cara Tate **SR. CONTENT PRODUCTION EDITORS** Harry Jach, Trista Wagoner **CONTENT PRODUCTION EDITORS** Jeffrey E. Cook, Chris Filiatreau, Cynthia Howe, Lauren Kmec, Barbara P. Ordway **SR. EDITORIAL COORDINATORS** Carolyn Kyle, Beverly Shields **EDITORIAL COORDINATORS** Ramatoulaye Diop, Joi S. Granger, Lisa Johnson, Anita Wynn **PUBLICATIONS ASSISTANTS** Aneera Dobbins, Jeffrey Hearn, Dona Mathieu, Le-Toya Mayne Flood, Shannon McMahon, Scott Miller, Jerry Richardson, Rachel Roberts(UK), Alice Whaley(UK), Brian White **EXECUTIVE ASSISTANT** Anna Bashkirova **ADMINISTRATIVE SUPPORT** Janet Clements(UK), Michael Crabtree(UK, Intern), Lizanne Newton(UK), Maryrose Madrid, John Wood(UK)

News

NEWS MANAGING EDITOR John Travis **INTERNATIONAL EDITOR** Richard Stone **DEPUTY NEWS EDITORS** Daniel Clery(UK), Robert Coontz, Elizabeth Culotta, David Grimm, David Malakoff, Leslie Roberts **CONTRIBUING EDITORS** Martin Enserink(Europe), Mara Hvistendahl (Asia) **SR. CORRESPONDENTS** Jeffrey Mervis, Elizabeth Pennisi **NEWS WRITERS** Adrian Cho, Jon Cohen, Jennifer Couzin-Frankel, Carolyn Gramling, Eric Hand, Jocelyn Kaiser, Kelly Servick, Robert F. Service, Erik Stokstad, Emily Underwood **INTERNS** David Shultz, Jia You **CONTRIBUTING CORRESPONDENTS** Pallava Bagla(South Asia), Michael Balter(Paris), John Bohannon, Ann Gibbons, Sam Kean, Richard A. Kerr, Eli Kintisch, Kai Kupferschmidt(Berlin), Andrew Lawler, Christina Larson(Beijing), Mitch Leslie, Charles C. Mann, Eliot Marshall, Virginia Morell, Dennis Normile(Tokyo), Heather Pringle, Tania Rabesandratana(Brussels), Gretchen Vogel(Berlin), Lizzie Wade(Mexico City) **CAREERS** Jim Austin(Editor), Donisha Adams **COPY EDITORS** Kara Estelle, Nora Kelly, Jennifer Levin **ADMINISTRATIVE SUPPORT** Scherraine Mack

Executive Publisher Alan I. Leshner
Publisher Kent R. Anderson **Chief Digital Media Officer** Rob Covey

BUSINESS OPERATIONS AND ADMINISTRATION DIRECTOR Deborah Rivera-Wienhold **BUSINESS SYSTEMS AND FINANCIAL ANALYSIS DIRECTOR** Randy Yi **MANAGER OF FULFILLMENT SYSTEMS** Neal Hawkins **SYSTEMS ANALYST** Nicole Mehmedovich **ASSISTANT DIRECTOR, BUSINESS OPERATIONS** Eric Knott **MANAGER, BUSINESS OPERATIONS** Jessica Tierney **BUSINESS ANALYSTS** Cory Lipman, Cooper Tilton, Celeste Troxler **FINANCIAL ANALYST** Jeremy Clay **RIGHTS AND PERMISSIONS ASSISTANT DIRECTOR** Emilie David **PERMISSIONS ASSOCIATE** Elizabeth Sandler **RIGHTS, CONTRACTS, AND LICENSING ASSOCIATE** Lili Kiser

MARKETING DIRECTOR Ian King **MARKETING MANAGER** Julianne Wielga **MARKETING ASSOCIATE** Elizabeth Sattler **SR. MARKETING EXECUTIVE** Jennifer Reeves **SR. ART ASSOCIATE, PROJECT MANAGER** Tzeitel Sorrosa **ART ASSOCIATE** Seil Lee **ASSISTANT COMMERCIAL EDITOR** Selby Frame **MARKETING PROJECT MANAGER** Angelissa McArthur **SR. WRITER** Bill Zimmer **PROGRAM DIRECTOR, AAAS MEMBER CENTRAL** Peggy Mihelich **FULFILLMENT SYSTEMS AND OPERATIONS** membership@aaas.org **MANAGER, MEMBER SERVICES** Pat Butler **SPECIALISTS** LaToya Casteel, Javia Flemmings, Latasha Russell **MANAGER, DATA ENTRY** Mickie Napoleoni **DATA ENTRY SPECIALISTS** JJ Regan, Jaimee Wise, Fiona Giblin

DIRECTOR, SITE LICENSING Tom Ryan **DIRECTOR, CORPORATE RELATIONS** Eileen Bernadette Moran **SR. PUBLISHER RELATIONS SPECIALIST** Kiki Forsythe **PUBLISHER RELATIONS MANAGER** Catherine Holland **PUBLISHER RELATIONS, EASTERN REGION** Keith Layson **PUBLISHER RELATIONS, WESTERN REGION** Ryan Rexroth **MANAGER, SITE LICENSE OPERATIONS** Iquo Edim **FULFILLMENT ANALYST** Lana Guz **ASSOCIATE DIRECTOR, MARKETING** Christina Schlecht **MARKETING ASSOCIATES** Thomas Landreth, Minah Kim

DIRECTOR OF WEB TECHNOLOGIES Ahmed Khadr **SR. DEVELOPER** Chris Coleman **DEVELOPERS** Dan Berger, Jimmy Marks **SR. PROJECT MANAGER** Trista Smith **SYSTEMS ENGINEER** Luke Johnson **PRODUCT MANAGER** Walter Jones

CREATIVE DIRECTOR, MULTIMEDIA Martyn Green **DIRECTOR OF ANALYTICS** Enrique Gonzales **SR. WEB PRODUCER** Sarah Crespi **WEB PRODUCER** Alison Crawford **VIDEO PRODUCER** Nguyen Nguyen **SOCIAL MEDIA PRODUCER** Meghna Sachdev

DIRECTOR OF OPERATIONS PRINT AND ONLINE Lizabeth Harman **PRINT PRODUCTION DIRECTOR** Wendy K. Shank **PREFLIGHT MANAGER** Marcus Spiegler **ASSISTANT MANAGER ONLINE** Lisa Stanford **ASSISTANT MANAGER PRINT** Rebecca Doshi **SR SPECIALISTS** Steve Forrester, Jason Hillman, Antoinette Hodal, Tara Kelly, Anthony Rosen **SPECIALISTS** Jacob Hedrick, Nichele Johnston, Lori Murphy, Kimberley Oster

DESIGN DIRECTOR Beth Rakouskas **ASSOCIATE ART DIRECTOR** Laura Creveling **SR. ILLUSTRATORS** Chris Bickel, Katharine Sutliff **ILLUSTRATOR** Valerie Altounian **SR. ART ASSOCIATES** Holly Bishop, Preston Huey **ART ASSOCIATES** Kay Engman, Garvin Grullón, Chrystal Smith **SR. PHOTO EDITOR** William Douthitt **PHOTO EDITOR** Leslie Blizard

DIRECTOR, GLOBAL COLLABORATION, CUSTOM PUBLICATIONS, ADVERTISING Bill Moran **EDITOR, CUSTOM PUBLISHING** Sean Sanders: 202-326-6430 **ASSISTANT EDITOR, CUSTOM PUBLISHING** Tianna Hicklin: 202-326-6463 **ADVERTISING MARKETING MANAGER** Justin Sawyers: 202-326-7061 science_advertising@aaas.org **ADVERTISING MARKETING ASSOCIATE** Javia Flemmings **ADVERTISING SUPPORT MANAGER** Karen Foote: 202-326-6740 **ADVERTISING PRODUCTION OPERATIONS MANAGER** Deborah Tompkins **SR. PRODUCTION SPECIALIST/GRAPHIC DESIGNER** Amy Hardcastle **PRODUCTION SPECIALIST** Yuse Lajiminmuhip **SR. TRAFFIC ASSOCIATE** Christine Hall **SALES COORDINATOR** Shirley Young **ASSOCIATE DIRECTOR, COLLABORATION, CUSTOM PUBLICATIONS/CHINA/TAIWAN/KOREA/SINGAPORE** Ruolei Wu: +86-186 0822 9345, rwu@aaas.org **COLLABORATION/CUSTOM PUBLICATIONS/JAPAN** Adarsh Sandhu + 81532-81-5142 asandhu@aaas.org **EAST COAST/E. CANADA** Laurie Faraday: 508-747-9395, FAX 617-507-8189 **WEST COAST/W. CANADA** Lynne Stickrod: 415-931-9782, FAX 415-520-6940 **MIDWEST** Jeffrey Dembksi: 847-498-4520 x3005, Steven Loerch: 847-498-4520 x3006 **UK EUROPE/ASIA** Roger Goncalves: TEL/FAX +41 43 243 1358 **JAPAN** Katsuyoshi Fukamizu(Tokyo): +81-3-3219-5777 kfukamizu@aaas.org **CHINA/TAIWAN** Ruolei Wu: +186-0082-9345

WORLDWIDE ASSOCIATE DIRECTOR OF SCIENCE CAREERS Tracy Holmes: +44 (0) 1223 326525, FAX +44 (0) 1223 326532 tholmes@science-int.co.uk **CLASSIFIED** advertise@sciencecareers.org **U.S. SALES** Tina Burks: 202-326-6577, Nancy Toema: 202-326-6578 **SALES ADMINISTRATOR** Marci Gallun **EUROPE/ROW SALES** Axel Gesatzki, Sarah Lelarge **SALES ASSISTANT** Kelly Grace **JAPAN** Hiroyuki Mashiki(Kyoto): +81-75-823-1109 hmashiki@aaas.org **CHINA/TAIWAN** Ruolei Wu: +86-186 0082 9345 rwu@aaas.org **MARKETING MANAGER** Allison Pritchard **MARKETING ASSOCIATE** Aimee Aponte

AAAS BOARD OF DIRECTORS RETIRING PRESIDENT, CHAIR Phillip A. Sharp **PRESIDENT** Gerald R. Fink **PRESIDENT-ELECT** Geraldine (Geri) Richmond **TREASURER** David Evans Shaw **CHIEF EXECUTIVE OFFICER** Alan I. Leshner **BOARD** Bonnie L. Bassler, May R. Berenbaum, Carlos J. Bustamante, Claire M. Fraser, Laura H. Greene, Elizabeth Loftus, Raymond Orbach, Inder M. Verma

SUBSCRIPTION SERVICES For change of address, missing issues, new orders and renewals, and payment questions: 866-434-AAAS (2227) or 202-326-6417, FAX 202-842-1065. Mailing addresses: AAAS, P.O. Box 96178, Washington, DC 20090-6178 or AAAS Member Services, 1200 New York Avenue, NW, Washington, DC 20005

INSTITUTIONAL SITE LICENSES 202-326-6755 **REPRINTS:** Author Inquiries 800-635-7181 **COMMERCIAL INQUIRIES** 803-359-4578 **PERMISSIONS** 202-326-6765, permissions@aaas.org **AAAS Member Services** 202-326-6417 or http://membercentral.aaas.org/discounts

Science serves as a forum for discussion of important issues related to the advancement of science by publishing material on which a consensus has been reached as well as including the presentation of minority of conflicting points of view. Accordingly, all articles published in *Science*—including editorials, news and comment, and books reviews—are signed and reflect the individual views of the authors and not official points of view adopted by AAAS or the institutions with which the authors are affiliated.

INFORMATION FOR AUTHORS See pages 680 and 681 of the 7 February 2014 issue or access www.sciencemag.org/about/authors

SENIOR EDITORIAL BOARD
A. Paul Alivisatos, *Lawrence Berkeley Nat'l. Laboratory,* **Ernst Fehr,** *U. of Zürich*
Susan M. Rosenberg, *Baylor College of Medicine,* **Michael S. Turner,** *U. of Chicago*

BOARD OF REVIEWING EDITORS (Statistics board members indicated with **(S)**)
Adriano Aguzzi, *U. Hospital Zürich*
Takuzo Aida, *U. of Tokyo*
Leslie Aiello, *Wenner-Gren Foundation*
Judith Allen, *U. of Edinburgh*
Sonia Altizer, *U. of Georgia*
Virginia Armbrust, *U. of Washington*
Sebastian Amigorena, *Institut Curie*
Kathryn Anderson, *Memorial Sloan-Kettering Cancer Center*
Peter Andolfatto, *Princeton U.*
Meinrat O. Andreae, *Max-Planck Inst. Mainz*
Paola Arlotta, *Harvard U.*
Johan Auwerx, *EPFL*
David Awschalom, *U. of Chicago*
Jordi Bascompte, *Estación Biológica de Doñana CSIC*
Facundo Batista, *London Research Inst.*
Ray H. Baughman, *U. of Texas, Dallas*
David Baum, *U. of Wisconsin*
Kamran Behnia, *ESPCI-ParisTech*
Yasmine Belkaid, *NIAID, NIH*
Philip Benfey, *Duke U.*
Stephen J. Benkovic, *Penn State U.*
Carlo Beenakker, *Leiden U.*
Gabriele Bergers, *U. of California, San Francisco*
Christophe Bernard, *Aix-Marseille U.*
Bradley Bernstein, *Massachusettes General Hospital*
Peer Bork, *EMBL*
Bernard Bourdon, *Ecole Normale Supérieure de Lyon*
Chris Bowler, *École Normale Supérieure*
Ian Boyd, *U. of St. Andrews*
Emily Brodsky, *U. of California, Santa Cruz*
Ron Brookmeyer, *U. of California Los Angeles* **(S)**
Christian Büchel, *U. Hamburg-Eppendorf*
Joseph A. Burns, *Cornell U.*
Gyorgy Buzsaki, *New York U. School of Medicine*
Blanche Capel, *Duke U.*
Mats Carlsson, *U. of Oslo*
David Clapham, *Children's Hospital Boston*
David Clary, *U. of Oxford*
Joel Cohen, *Rockefeller U., Columbia U.*
Jonathan D. Cohen, *Princeton U.*
James Collins, *Boston U.*
Robert Cook-Deegan, *Duke U.*
Alan Cowman, *Walter & Eliza Hall Inst.*
Robert H. Crabtree, *Yale U.*
Roberta Croce, *Vrije Universiteit*
Janet Currie, *Princeton U.*
Jeff L. Dangl, *U. of North Carolina*
Tom Daniel, *U. of Washington*
Frans de Waal, *Emory U.*
Stanislas Dehaene, *Collège de France*
Robert Desimone, *MIT*
Claude Desplan, *New York U.*
Ap Dijksterhuis, *Radboud U. of Nijmegen*
Dennis Discher, *U. of Pennsylvania*
Gerald W. Dorn II, *Washington U. School of Medicine*
Jennifer A. Doudna, *U. of California, Berkeley*
Bruce Dunn, *U. of California, Los Angeles*
Christopher Dye, *WHO*
Todd Ehlers, *U. of Tuebingen*
David Ehrhardt, *Carnegie Inst. of Washington*
Tim Elston, *U. of North Carolina at Chapel Hill*
Gerhard Ertl, *Fritz-Haber-Institut, Berlin*
Barry Everitt, *U. of Cambridge*
Ernst Fehr, *U. of Zurich*
Anne C. Ferguson-Smith, *U. of Cambridge*
Michael Feuer, *The George Washington U.*
Kate Fitzgerald, *U. of Massachusetts*
Peter Fratzl, *Max-Planck Inst.*
Elaine Fuchs, *Rockefeller U.*
Daniel Geschwind, *UCLA*
Andrew Gewirth, *U. of Illinois*
Karl-Heinz Glassmeier, *TU Braunschweig*
Ramon Gonzalez, *Rice U.*
Julia R. Greer, *Caltech*
Elizabeth Grove, *U. of Chicago*
Kip Guy, *St. Jude's Children's Research Hospital*
Taekjip Ha, *U. of Illinois at Urbana-Champaign*
Christian Haass, *Ludwig Maximilians U.*
Steven Hahn, *Fred Hutchinson Cancer Research Center*
Michael Hasselmo, *Boston U.*
Martin Heimann, *Max-Planck Inst. Jena*
Yka Helariutta, *U. of Cambridge*
James A. Hendler, *Rensselaer Polytechnic Inst.*
Janet G. Hering, *Swiss Fed. Inst. of Aquatic Science & Technology*
Michael E. Himmel, *National Renewable Energy Lab.*
Kai-Uwe Hinrichs, *U. of Bremen*
Kei Hirose, *Tokyo Inst. of Technology*
David Hodell, *U. of Cambridge*
David Holden, *Imperial College*
Lora Hooper, *UT Southwestern Medical Ctr. at Dallas*
Raymond Huey, *U. of Washington*
Steven Jacobsen, *U. of California, Los Angeles*
Kai Johnsson, *EPFL Lausanne*
Peter Jonas, *Inst. of Science & Technology (IST) Austria*
Matt Kaeberlein, *U. of Washington*
William Kaelin Jr., *Dana-Farber Cancer Inst.*
Daniel Kahne, *Harvard U.*
Daniel Kammen, *U. of California, Berkeley*
Masashi Kawasaki, *U. of Tokyo*
Joel Kingsolver, *U. of North Carolina at Chapel Hill*
Robert Kingston, *Harvard Medical School*
Eitonne Koechlin, *Ecole Normale Suprieure*
Alexander Kolodkin, *Johns Hopkins U.*
Roberto Kolter, *Harvard Medical School*
Alberto R. Kornblihtt, *U. of Buenos Aires*
Leonid Kruglyak, *UCLA*
Thomas Langer, *U. of Cologne*
Mitchell A. Lazar, *U. of Pennsylvania*
David Lazer, *Harvard U.*
Thomas Lecuit, *IBDM*
Virginia Lee, *U. of Pennsylvania*
Stanley Lemon, *U. of North Carolina at Chapel Hill*
Ottoline Leyser, *Cambridge U.*
Marcia C. Linn, *U. of California, Berkeley*
Jianguo Liu, *Michigan State U.*
Luis Liz-Marzan, *CIC biomaGUNE*
Jonathan Losos, *Harvard U.*
Ke Lu, *Chinese Acad. of Sciences*
Christian Lüscher, *U. of Geneva*
Laura Machesky, *CRUK Beatson Inst. for Cancer Research*
Anne Magurran, *U. of St. Andrews*
Oscar Marin, *CSIC & U. Miguel Hernández*
Charles Marshall, *U. of California, Berkeley*
C. Robertson McClung, *Dartmouth College*
Graham Medley, *U. of Warwick*
Yasushi Miyashita, *U. of Tokyo*
Richard Morris, *U. of Edinburgh*
Alison Motsinger-Reif, *NC State U.* **(S)**
Sean Munro, *MRC Lab. of Molecular Biology*
Thomas Murray, *The Hastings Center*
James Nelson, *Stanford U. School of Med.*
Karen Nelson, *J. Craig Venter Institute*
Daniel Neumark, *U. of California, Berkeley*
Timothy W. Nilsen, *Case Western Reserve U.*
Pär Nordlund, *Karolinska Inst.*
Helga Nowotny, *European Research Advisory Board*
Ben Olken, *MIT*
Joe Orenstein, *U. of California Berkeley & Lawrence Berkeley National Lab*
Harry Orr, *U. of Minnesota*
Andrew Oswald, *U. of Warwick*
Steve Palumbi, *Stanford U.*
Jane Parker, *Max-Planck Inst. of Plant Breeding Research*
Giovanni Parmigiani, *Dana-Farber Cancer Inst.* **(S)**
Donald R. Paul, *U. of Texas, Austin*
John H. J. Petrini, *Memorial Sloan-Kettering Cancer Center*
Joshua Plotkin, *U. of Pennsylvania*
Albert Polman, *FOM Institute AMOLF*
Philippe Poulin, *CNRS*
David Randall, *Colorado State U.*
Colin Renfrew, *U. of Cambridge*
Felix Rey, *Institut Pasteur*
Trevor Robbins, *U. of Cambridge*
Jim Roberts, *Fred Hutchinson Cancer Research Ctr.*
Barbara A. Romanowicz, *U. of California, Berkeley*
Jens Rostrup-Nielsen, *Haldor Topsoe*
Mike Ryan, *U. of Texas, Austin*
Mitinori Saitou, *Kyoto U.*
Shimon Sakaguchi, *Kyoto U.*
Miquel Salmeron, *Lawrence Berkeley National Lab*
Jürgen Sandkühler, *Medical U. of Vienna*
Alexander Schier, *Harvard U.*
Randy Seeley, *U. of Cincinnati*
Vladimir Shalaev, *Purdue U.*
Robert Siliciano, *Johns Hopkins School of Medicine*
Joseph Silk, *Institut d'Astrophysique de Paris*
Denis Simon, *Arizona State U.*
Alison Smith, *John Innes Centre*
Richard Smith, *U. of North Carolina* **(S)**
John Speakman, *U. of Aberdeen*
Allan C. Spradling, *Carnegie Institution of Washington*
Jonathan Sprent, *Garvan Inst. of Medical Research*
Eric Steig, *U. of Washington*
Paula Stephan, *Georgia State U. and National Bureau of Economic Research*
Molly Stevens, *Imperial College London*
V. S. Subrahmanian, *U. of Maryland*
Ira Tabas, *Columbia U.*
Sarah Teichmann, *Cambridge U.*
John Thomas, *North Carolina State U.*
Shubha Tole, *Tata Institute of Fundamental Research*
Christopher Tyler-Smith, *The Wellcome Trust Sanger Inst.*
Herbert Virgin, *Washington U.*
Bert Vogelstein, *Johns Hopkins U.*
Cynthia Volkert, *U. of Göttingen*
Douglas Wallace, *Dalhousie U.*
David Wallach, *Weizmann Inst. of Science*
Ian Walmsley, *U. of Oxford*
David A. Wardle, *Swedish U. of Agric. Sciences*
David Waxman, *Fudan U.*
Jonathan Weissman, *U. of California, San Francisco*
Chris Wikle, *U. of Missouri* **(S)**
Ian A. Wilson, *The Scripps Res. Inst.* **(S)**
Timothy D. Wilson, *U. of Virginia*
Rosemary Wyse, *Johns Hopkins U.*
Jan Zaanen, *Leiden U.*
Kenneth Zaret, *U. of Pennsylvania School of Medicine*
Jonathan Zehr, *U. of California, Santa Cruz*
Len Zon, *Children's Hospital Boston*
Maria Zuber, *MIT*

BOOK REVIEW BOARD
David Bloom, *Harvard U.,* **Samuel Bowring,** *MIT,* **Angela Creager,** *Princeton U.,* **Richard Shweder,** *U. of Chicago,* **Ed Wasserman,** *DuPont*

Published by AAAS

EDITORIAL

Zero infection

Last month, United Nations (UN) Secretary-General Ban Ki-moon, World Bank Group President Jim Yong Kim, and World Health Organization (WHO) Director-General Margaret Chan held a press conference after the UN Chief Executive Board's meeting to assess the global response to the Ebola epidemic. The secretary-general's message was clear: "Our end game is not near." To eliminate infection from the human population by mid-2015, as Ban Ki-moon hopes, the world must intensify its fight against this virus, but we should also recognize that we need better ways to combat international health hazards of all kinds.

One year into West Africa's Ebola epidemic, international attention remains sharply focused on strengthening the response in Guinea, Liberia, and Sierra Leone, where hundreds of new cases arise each week. Its recent spread into Mali is another warning that if epidemics such as Ebola are to be contained, then they must be tackled at their source; that means rapid detection and effective action at the earliest, more manageable stages. To better achieve this in future, we need to bolster global surveillance and enhance national and international capability to react appropriately.



"...we need to bolster global surveillance and enhance ...capability to react appropriately."

The foundations of a more effective response already exist in the form of multidisciplinary technical networks, WHO's Global Outbreak and Alert Response Network, the Training Programs in Epidemiology and Public Health Interventions Network, the logistics and special competences of the international humanitarian system, and the International Health Regulations. But these existing elements need to be brought into a stronger system of global governance that can manage a variety of health hazards, especially infectious diseases. A new mechanism could be steered by a single international agency, possibly an existing organization. Irrespective of who leads, multiple agencies must participate: WHO, the World Bank, the World Organization for Animal Health, the World Food Programme, and the International Monetary Fund, among others.

Existing efforts need to be reinforced and placed within a larger, more cohesive operational framework. That framework would provide more-predictable access to expertise, funding, and training, and on a large scale. It would recognize that most emerging infections, including Ebola, have an animal origin, so the veterinary sector would have a key role. Building national capacity is paramount, and the strengthening of national health services is vital. The new mechanism would include an adequately resourced global rapid reaction capability to manage major incidents and provide immediate and sustained assistance where local health services are at risk of being overwhelmed, or where there is a major risk of international spread.

Ebola has underlined the need, and presented an opportunity, to modernize outbreak management. Traditionally, the detection and monitoring of outbreaks have relied on a paper trail of clinical case reports. This is typically insufficient, but can now be supplemented by other techniques, such as the use of handheld data-collection devices, social media, serosurveillance, risk modeling, and analysis of pathogen genome data. Gathering such information is sometimes seen as an unnecessary distraction from immediate health needs. Yet these data are integral to the successful management of outbreaks today.

Preparedness and readiness reflect what happens before, rather than during, a crisis, irrespective of the hazard. We need to improve contingency planning and coordination; develop and stockpile diagnostics, drugs, and vaccines; set up sequencing pipelines and data-sharing protocols; and anticipate public engagement and ethical issues. We need to recognize that disease surveillance and response are global public goods. Ebola reminds us that wherever new diseases emerge, they can quickly and easily become international problems. In the event of a crisis, prior investment in infrastructure, training, and research will pay off many times over.

We can drive Ebola infection out of the human population, but we must also do more to guard against, and respond to, similar crises in the future. All options for doing that are now on the table.

– **Mark Woolhouse, Patrick Drury, Christopher Dye**

Mark Woolhouse is a professor at the Centre for Immunity, Infection & Evolution at the University of Edinburgh, Edinburgh, UK.

Patrick Drury is Manager of the Global Outbreak Alert and Response Network of the World Health Organization, Geneva, Switzerland.

Christopher Dye is the Director of Strategy in the Office of the Director General at the World Health Organization, Geneva, Switzerland. E-mail: dyec@who.int

10.1126/science.aaa4117

ILLUSTRATION: NOPPARIT/ISTOCKPHOTO.COM

Downloaded from www.sciencemag.org on December 14, 2014

Published by AAAS

NEWS

> "I look at some of the titles and I go—really?"
>
> **University of California President Janet Napolitano**, on the need to improve federal grant descriptions, a step the National Science Foundation plans to take to eliminate fodder for congressional critics (http://scim.ag/NSFgrants).

IN BRIEF



Artist's rendition of Hayabusa 2 on its way to asteroid 1999 JU3.

Aiming for the heart of an asteroid

Japan last week successfully launched the spacecraft Hayabusa 2 on a 6-year round trip to asteroid 1999 JU3. If all goes well, the spacecraft will return to Earth in 2020 with the first-ever samples from beneath an asteroid's surface. The probe will circle the asteroid, collecting data, deploying several small landers, and briefly touching down to collect surface samples. Then, in the most audacious gambit of the mission, Hayabusa 2 will release an impactor that will blast a crater into the asteroid. The spacecraft will touch down a second time and capture ejecta. Comparing surface and underground samples will give scientists a better idea of the original material of this asteroid type and how space weathers it. The Hayabusa 2 team hopes its probe will have a less eventful voyage than its predecessor: The first Hayabusa overcame fuel leaks, equipment failures, and a weeks-long loss of communications during its mission to the asteroid Itokawa in 2005. But Hayabusa set a historic first: In 2010, it carried samples from an asteroid's surface to Earth. http://scim.ag/Haya2launch

AROUND THE WORLD

Chimp 'personhood' appeal fails

ALBANY | Advocates of "legal personhood" to chimpanzees have lost another battle. Last week, a New York appellate court rejected a lawsuit by the Nonhuman Rights Project (NhRP) to free a chimp named Tommy from captivity. The group had argued that the chimpanzee deserved the human right of bodily liberty. The court's decision is the latest setback for NhRP, an animal rights group that has been trying to free four New York chimpanzees—including two research chimps—since 2013 (*Science*, 6 December 2013, p. 1154). Despite the loss, the group is pursuing more cases in the hopes of conferring legal rights to a variety of animals, from elephants to dolphins. http://scim.ag/chimppers

New cell science institute

PHILADELPHIA, PENNSYLVANIA | Eleven years after establishing the Allen Institute for Brain Science, Microsoft co-founder Paul Allen has a sequel: the Allen Institute for Cell Science. Like its predecessor, the new institute will be seeded with $100 million, and the data generated will be publicly available. (It will also be housed in the same Seattle building.) It aims to decipher how the cell's components work together and how they are perturbed by gene mutations, drugs, and other forces. Rick Horwitz, formerly a professor of cell biology at the University of Virginia, will be executive director; initial plans call for hiring 75 scientists. The first project, announced 8 December at the annual meeting of the American Society for Cell Biology, will focus on how induced pluripotent stem cells transform into heart muscle and epithelial cells. http://scim.ag/AllenCell

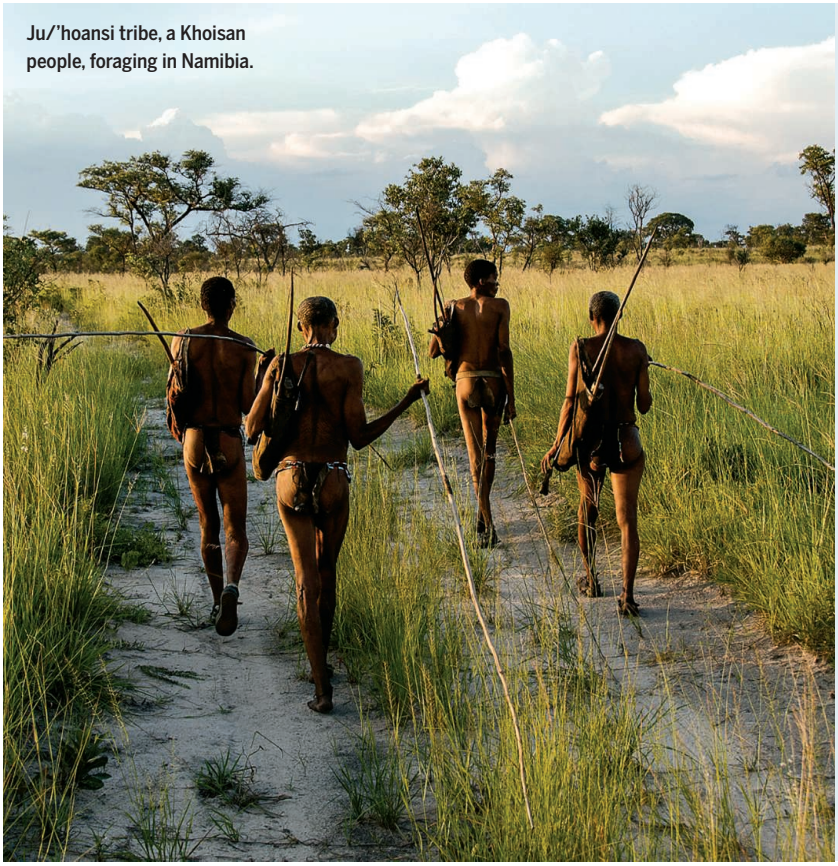
Colorful tombs on the Silk Road

KUCHA, CHINA | In the rubble of the ancient Silk Road, Chinese archaeologists have excavated the ruins of tombs dating to the 3rd or 4th century C.E. Discovered in 2007 in northwest China during one of the country's relentless construction projects, the tombs give insight into the ways of the

IMAGE: JAXA/AKIHIRO IKESHITA

Downloaded from www.sciencemag.org on December 14, 2014

Published by AAAS



Ju/'hoansi tribe, a Khoisan people, foraging in Namibia.

Shrinking tribe once most populous

The Kalahari Bushmen of southern Africa—who speak in Khoisan "click" languages—are in decline, with only 100,000 speakers in Africa. But the Khoisan have the greatest genetic diversity known among human populations, despite their small size. Based on a single genome published in 2012, researchers thought the Khoisan came from a large ancestral population but couldn't rule out interbreeding with other diverse Africans. The new study, in *Nature Communications*, compared genomes from five Khoisan hunter-gatherers with DNA from 1462 genomes of other people—and found that two of the bushmen inherited their DNA only from northern Khoisan ancestors whose lineages were very ancient. The team also found that the populations of four groups—ancestors of the Khoisan, the Yoruba, Europeans, and Asians—all declined 120,000 to 30,000 years ago, likely due to dry climate. But the Khoisan decreased far less than the others—possibly because of increases in rainfall in southern Africa that supported their population. That suggests that for tens of thousands of years, the Khoisan were "the largest population" on Earth. **http://scim.ag/Kalatribe**

wealthy in Qiuci, then a powerful oasis state in the Taklamakan desert. Yu Zhiyong and colleagues from the Xinjiang Institute of Cultural Relics and Archaeology excavated 10 tombs, publishing their findings in October in *Chinese Cultural Relics*. The tombs were colorful and intricate: Blue brick gates opened into domed burial chambers lit by oil lamps. In one chamber, the complete skeletons of four women and one man were found face-up; coins nearby suggest two may have held money. The chambers appear to have been used for repeated multiple burials, the archaeologists wrote, warranting further research.



Tomb M1, one of 10 ancient tombs unearthed in the oasis state of Qiuci in 2007.

PHOTOS: (TOP TO BOTTOM) STEPHAN C. SCHUSTER, PENNSYLVANIA STATE UNIVERSITY; CHINESE CULTURAL RELICS

2014 may be hottest on record

LIMA | The World Meteorological Organization (WMO) announced on 3 December at the U.N. Framework Convention on Climate Change talks that 2014 is on track to be the warmest year on record. Whether the 2014 record, if confirmed, means the end of the so-called warming hiatus is unclear, WMO's Michel Jarraud said during a teleconference. The hiatus, which began in about 1998, has been typified by a slower rate of increase in average air temperatures than in previous decades. **http://scim.ag/2014warmest**

Cape Cod turtle deaths

QUINCY, MASSACHUSETTS | Hundreds of endangered sea turtles, mostly juvenile Kemp's ridleys less than a foot long, have been washing up on the beaches of Cape Cod, Massachusetts, sick and stunned by the cold ocean water. They are being trapped on their southbound fall migration to warmer climes by the arm of the cape, which protrudes into the Atlantic Ocean. Many also have life-threatening conditions like dehydration, pneumonia, infections, or off-kilter blood chemistry. Rescue crews have picked up more than 1070 turtles, about 20% of them already dead; the living turtles are sent to a sea turtle hospital in Quincy, Massachusetts, run by the New England Aquarium. Six hundred and fifty turtles have been admitted so far—approaching triple the hospital's previous record of 240, set in 2012. The reasons for this year's stranding remain unknown. **http://scim.ag/turtledeaths**

Habitat proposed for Arctic seals

WASHINGTON, D.C. | Shrinking sea ice is wreaking havoc on Arctic ringed seals, which live in and beneath the ice in the Bering, Chukchi, and Beaufort seas and were listed as threatened under the Endangered Species Act in December 2012. Last week, the National Oceanic and Atmospheric Administration (NOAA) proposed designating more than 906,000 square kilometers off Alaska as protected habitat for the seals, based on a yearlong study of their critical habitat needs and the economic effects that protecting them may have on the region, which harbors extensive oil and gas reserves

Published by AAAS



A ringed seal pup on an iceberg in Nunavut Territory, Canada.

and commercially valuable fish. NOAA will accept public comments on the proposal until 3 March 2015. **http://scim.ag/sealhabitat**

ArXiv probes 'text reuse'

ITHACA, NEW YORK | Two analyses of manuscripts submitted to arXiv, the world's largest preprint repository, examine the prevalence of scientific plagiarism. One study of "text reuse" in 757,000 articles submitted between 1991 and 2012 found that that the more text a paper poached from published work, the less frequently it was cited. The study was published 8 December in the *Proceedings of National Academy of Sciences*. A second analysis by Science's news staff found that 3.2% of 301,759 arXiv papers submitted since 2011 were "flagged" by repository software for extensive text reuse without citation. Of the 57 nations with more than 100 submitting authors, Bulgaria had the highest rate of flagged authors, 20%; New Zealand had the lowest, 0.6%. **http://scim.ag/_textreuse**

Renewal again for R&D tax break?

WASHINGTON, D.C. | The U.S. Congress was set this week to renew a popular tax break that allows companies to write off certain research expenses. But industry and a bipartisan group of lawmakers failed in their bid to make the R&D tax credit, worth some $7 billion annually, a permanent part of the tax code—as many economic experts have urged. Instead, as *Science* went to press, the Senate was poised to pass a House of Representatives–approved bill that would temporarily restore for 2014 the R&D credit and nearly 50 other tax breaks that expired earlier this year. It would mark the 16th time Congress has temporarily extended the credit since it was created in 1981. Critics say that pattern makes it difficult for businesses to plan long-term investments. **http://scim.ag/RandDtaxcred**

BY THE NUMBERS

$4.1 million

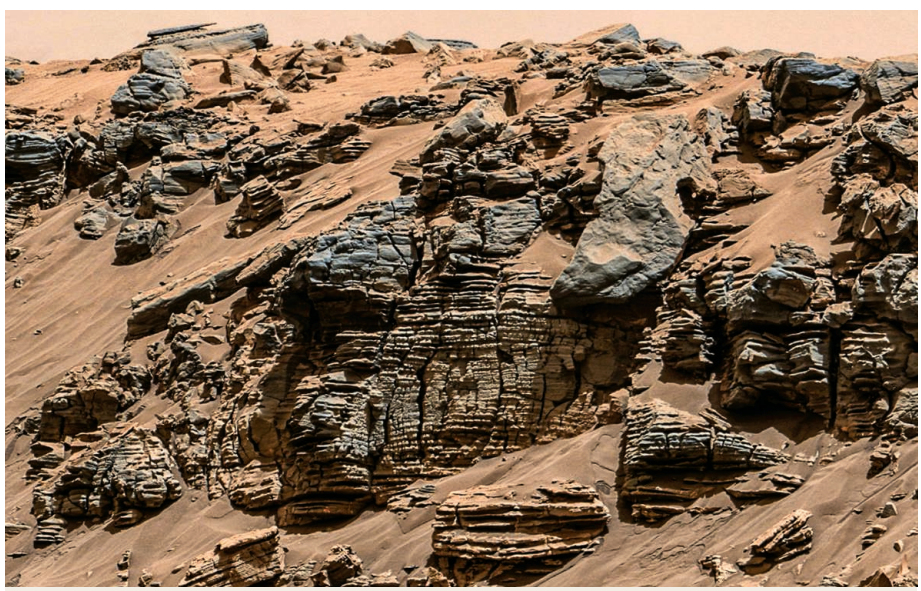
Sale price for James Watson's Nobel Prize, earned in 1962 for the co-discovery of the structure of DNA and auctioned off last week. The buyer, Russian mogul Alisher Usmanov, has said he plans to return it to Watson.

269,000

Total weight, in tons, of plastic floating in the world's oceans, according to a new study in *PLOS ONE*.

8 million

Number of "dark" asteroids that may be lurking in the Oort cloud, scientists reported last month in the *Monthly Notices of the Royal Astronomical Society*.



Lakebeds in Mars's Gale crater were built over millions of years.

Mars crater held ancient lake

Gale crater, the bowl on Mars that NASA's Curiosity rover has been exploring for 2.5 years, was filled with water billions of years ago—and had bouts of deposition in river deltas and lakebeds that may have lasted a million years or more, scientists reported 8 December in a press conference. That's good news for astrobiologists who hope that the ancient environment on the Red Planet was habitable for a sustained time. Mars scientists always suspected that Gale crater held water; in December 2013, the Curiosity team reported evidence for lakebed mudstones near the rover's landing spot. This week, the team described angled sediment beds associated with river delta deposition and the flat bedding of lake sediments, suggesting million-year cycles of deposition. As Curiosity explores a sediment mountain in the crater's center, it may spy other deposition cycles, which could boost estimates for the duration of the wet period to tens of millions of years. **http://scim.ag/_MarsGale**

PHOTOS: (TOP TO BOTTOM) © PAUL SOUDERS/CORBIS; NASA/JPL-CALTECH/MSSS

Published by AAAS







GENOMICS

Bird genomes give new perches to old friends

Comparing genomes clarifies family relations and pinpoints genes for song learning

By **Elizabeth Pennisi**

With chicks that have claws on their wings and a digestive system that resembles a cow's, the pheasant-sized hoatzin that roams the Amazon has always puzzled those trying to place it within the avian family tree. But now researchers believe they have pinned down the odd bird's relatives—just one of the many findings revealed this week from a massive international project analyzing the sequenced genomes of 48 bird species representing nearly every order of bird. The fruits of this effort—eight papers this week in *Science* and more than 20 additional reports in several other journals—represent the biggest advance in avian biology in decades. "This has not been done for any other organism before," says Per Ericson, an evolutionary biologist at the Swedish Museum of Natural History in Stockholm. "It's mind-blowing."

The effort, involving 200 people from 80 labs and several months of supercomputer time, has yielded the most definitive avian family tree yet. It has also pinpointed gene networks underlying the traits that make birds birds, such as feathers and beaks instead of teeth. In one provocative finding, a team has identified the gene network that underlies complex singing in birds—and found that the same network operates in humans, where it is presumably crucial to language.

Curiosity about vocal learning was what led Erich Jarvis to coordinate the project in the first place. Jarvis, who is a neurobiologist at Duke University in Durham, North Carolina, wanted to know how many times the ability to imitate sounds had evolved in birds. But he didn't trust the existing family trees, which were built on comparisons of less than a dozen genes or several thousand bases.

Jarvis began thinking that the solution was a tree based on whole genomes—which provide millions of bases for comparison. And at the same time, other researchers had begun an ambitious effort to sequence 10,000 vertebrate species, and China's sequencing giant, BGI, had agreed to kick that off by tackling 101 species, including 11 birds. In 2010, Jarvis and Tom Gilbert, a paleogenomicist at the University of Copenhagen, helped convince BGI and its collaborators to increase the number of bird species they did to 43. By the end of 2011, sequencing of 48 in total was done.

That proved the easy part. When the researchers tried to build the new avian family tree, "we were shocked to find we couldn't get a solid answer," Jarvis recalls. As the consortium developed more sophisticated bioinformatics tools to analyze the genome data, they discovered that protein-coding genes by themselves were not the most reliable for building good trees. The noncoding regions within or between genes, called introns, gave better answers. And although the group had access to supercomputers, they still had to come up with a way to allocate the analysis to the machines' many microprocessors. "It took 3 years to iron out the kinks," Gilbert says.

The new avian tree, described on page 1320 by Jarvis's team, resembles one built in 2008 by others, based only on introns from 169 species. Some of the branches in that earlier tree were heretical at the time, but "a lot of the same relationships hold up even when you have access to entire genomes," says Sushma Reddy, an evolutionary biologist at Loyola University Chicago in Illinois who took part in the earlier analysis. For example, the 2008 paper indicated that doves, sandgrouses, and flightless Madagascan

With 48 bird genomes in hand, researchers discovered that hummingbirds (*top*), budgerigars (*right*), and zebra finches (*bottom*) sing using the same network of genes as humans talking.

PHOTOS: (TOP) PAT GAINES/FLICKR; (BOTTOM) YOUGETTHEPICTURE/FLICKR; (RIGHT) TIM GRAHAM/GETTY IMAGES

Downloaded from www.sciencemag.org on December 14, 2014

Published by AAAS

birds called mesites are all closely related and resemble the ancestors of so-called modern birds, which include all existing species except for chickens, ducks, and flightless birds such as the ostrich. Likewise, in 2008, few accepted a claim that flamingos and grebes are close relatives, but the new work confirms the link.

In contrast with the 2008 tree, the new analysis identifies the doves, sandgrouses, and mesites as close relatives of flamingos and grebes, creating a different view of the early modern birds. "That's the key phylogenic finding," says co-author Joel Cracraft, an evolutionary biologist at the American Museum of Natural History in New York City. And it reveals that the hoatzin is not a relative of the cuckoos—which it faintly resembles—but instead belongs with cranes and plovers.

The results go beyond family relations, showing, for example, that song learning evolved three separate times: on the branches leading to songbirds, hummingbirds, and parrots. And on page 1311, Guojie Zhang of BGI in Shenzhen, China, and his colleagues confirm that bird genomes are smaller than other vertebrate genomes. They have less repetitive DNA and have also lost 7% of the genes—58 million bases in total—found in a lizard cousin, the anole. Comparisons with other nonavian species revealed bird-specific genes, such as those for keratin proteins likely involved in feather development, Zhang says. "We can take an evolutionary view of the genome the way we have an evolutionary view of anatomy," says Jarvis co-author Stephen O'Brien, a geneticist at St. Petersburg State University in Russia.

Take Jarvis's vocal learning question. For decades, researchers have known that human speech and avian song learning are similar behaviors. Jarvis and others have since shown that songbirds, parrots, hummingbirds, and humans use equivalent forebrain regions for these processes as well. One study even showed that both feathered and nonfeathered vocal learners rely on a gene called *FOXP2*. The new avian genomes have now enabled Jarvis, Andreas Pfenning of the Massachusetts Institute of Technology in Cambridge, and their colleagues to assess whether the same gene networks always underlie vocal learning.

Pfenning first compared patterns of gene activity in different parts of the brains of the parakeet, hummingbird, and zebra finch, which represent the three avian lineages in which imitative singing arose. They also studied two birds that lack vocal learning, the ring dove and quail. The same pattern emerged in all three song-learning birds, but not in the quail and the dove.

The researchers went on to compare the song-related pattern with other researchers' maps of gene activity in the brains of humans and of macaques—which don't have vocal communication as complex as human speech. Similar genes were at work in the human brain regions associated with speech, they report on page 1333, while the pattern was missing in the macaque. "It really nicely matches the behavioral and anatomical convergences," says Kelsey Martin, a neurobiologist at the University of California, Los Angeles (UCLA). To Martin, the gene network highlighted by the work resembles those that other animals use for spatial learning. "It suggests there's a core pattern of gene expression for learning," she says.

On page 1334, Jarvis's lab, led by Osceola Whitney, offers a closer look at gene activity patterns in the brains of adult zebra finches after they had been singing. The team identifies more than 2700 genes regulated by singing, with different sets of them active in each of the four song-related brain structures. "These two [papers] are both tours de force," says Stephanie White, a UCLA neuroscientist. With them, "the songbird may be getting its recognition for what it has to offer the speech and language community."

> ***"We're as close as we've ever been to the true phylogeny."***
>
> **Per Ericson**, Swedish Museum of Natural History

Though the avian genome project was ambitious, Jarvis and his colleagues view it as just a first step. Pfenning and Whitney, for example, now plan to probe the functions of the genes newly tied to vocal learning. And the bird family tree still needs refining.

In total, six of the 46 nodes, or branch points, in the new tree remain uncertain, Cracraft notes. One reason for the ambiguity is that birds, like mammals, underwent rapid diversification 66 million years ago, when the disappearance of most dinosaurs opened up many niches for new species to fill. The swift pace of evolution resulted in genomes that are a patchwork of novel and shared regions, making them hard to analyze and leaving some questions still unanswered—where owls belong, for example.

It will take more bird genomes to solve many of the puzzles, says Shannon Hackett, an ornithologist at the Field Museum in Chicago and lead author of the 2008 bird tree. "We're as close as we've ever been to the true phylogeny," Ericson says. "[But] we are not there yet." That may not take long: At BGI, the genomes of about 200 more birds are done and waiting to be analyzed, with thousands more planned. ■

INFECTIOUS DISEASES

The genetics of resistant malaria

Studies confirm changes in suspected gene confer resistance to artemisinin

By **Gretchen Vogel**

The region around the Mekong River delta is infamous for its malaria parasites. Twice already—in the 1950s and the 1960s—they have developed resistance to key drugs, and the underlying mutations spread inexorably around the world, forcing public health officials to find new ways to fight the disease. Now it is happening again. Over the last decade, artemisinin, the most powerful drug available to cure malaria, has failed in more and more people in Cambodia, Myanmar, Vietnam, Laos, and border regions of Thailand. Researchers and public health experts worry that history will repeat itself and resistant parasites will go global. With any new drugs still years away from clinical use, that would be a disaster.

Money has poured into efforts to contain and eliminate the region's resistant strains, so far without success (*Science*, 14 May 2010, p. 844). This week, two papers online in *Science* offer new insights into the genes behind the threat. One helps explain which genetic changes allow the parasites to survive the drug (http://scim.ag/JStraimer). The other details how the mutations protect the parasites: by slowing their development and ramping up their defenses against the kinds of protein damage that artemisinin seems to cause (http://scim.ag/SMok). Such insights should help scientists identify and track resistant parasites and perhaps find better ways to kill them. The studies are "extremely interesting and important," says Pascal Ringwald, who coordinates the malaria drug resistance and containment program for the World Health Organization in Geneva, Switzerland.

Artemisinin is derived from sweet wormwood, a plant used for centuries in China to fight fevers. In the 1970s, Chinese researchers discovered that artemisinin made a powerful malaria drug. It, and several related derivatives, have since helped drive spectacular drops in the number of malaria deaths worldwide. The drug is easy to administer,

Published by AAAS



Children in Prey Mong kol village in Pailin province, Cambodia—a region where nearly three-quarters of malaria infections are resistant to artemisinin-based drugs—sleep under insecticide-treated bed nets.

has few side effects, and acts quickly—killing most parasites within hours. To prevent the emergence of resistance, the drug is paired with different partner drugs in so-called artemisinin-based combination therapies, or ACTs.

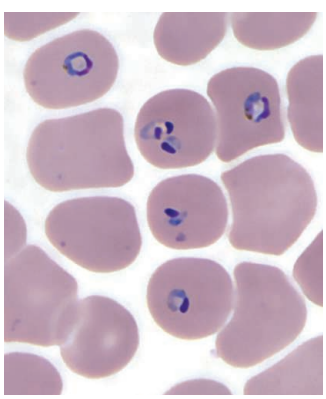
But in 2003 and 2004, the first signs of resistance to ACTs showed up on the Thai-Cambodia border. At first, many researchers didn't want to believe the artemisinin derivatives were failing; instead they blamed failures on the partner drug. But by 2009, it was clear that artemisinins were indeed losing efficacy. The race was on to find out what genes were behind the resistance—and eliminate the parasites that carry them.

A year ago, researchers reported in *Nature* that mutations in a gene for a protein called the K13 propeller were associated with artemisinin resistance in the deadly malaria parasite *Plasmodium falciparum*. (The protein is named for its pinwheel shape and its gene's location on chromosome 13 in the microbes.) And a July paper in *The New England Journal of Medicine* (*NEJM*) showed that K13 mutations were found in drug-resistant infections across Southeast Asia.

Now, David Fidock of Columbia University's College of Physicians and Surgeons and his colleagues have direct, laboratory evidence that the K13 gene is guilty as charged. "They nail the fact that the K13 mutations are indeed responsible for transferring artemisinin resistance," says Arjen Dondorp of the Mahidol Oxford Tropical Medicine Research Unit in Bangkok.

The incriminating evidence comes from a technique for precisely editing genes, called zinc finger nuclease technology. With it, Fidock and his colleagues could either repair the mutant K13 gene in resistant parasites or introduce a mutation in the gene in parasites still susceptible to artemisinin. They found that inserting a normal K13 gene into resistant parasites made them vulnerable to the drugs, and inserting certain K13 mutations into susceptible parasites made those resistant.

The second study suggests how K13 mutations exert their effects. In it, Zbynek Bozdech of Nanyang Technological University in Singapore and his colleagues, including Dondorp, analyzed the transcriptome—all the messenger RNA molecules—of parasites from more than 1000 patients from Southeast Asia and Africa. They found that parasites with K13 mutations tended to upregulate genes involved in protein folding and repair, while lowering expression of genes involved in DNA replication. That may help them repair the damage caused by artemisinin, which is thought to damage proteins by releasing highly reactive molecules.



Immature ring-form stage of malaria parasites in red blood cells.

The researchers also showed that resistant parasites spend more time in an immature form called the ring stage—temporarily slowing their development and buying time to recover before they try to mature to the next stage. The resistant parasites seem to be "ready to take a punch," says Rick Fairhurst of the National Institute of Allergy and Infectious Diseases in Rockville, Maryland, who is an author on both papers. "They have their protein repair and turnover pathways ready to go" when artemisinin hits.

Malaria researchers can take some solace in one finding, however. In the zinc finger study, some of the mutations introduced into K13 strongly protected the pathogen from the drug, whereas others had a weaker effect. Yet one of the mutations most commonly found in people with drug-resistant infections produces only weak resistance in the lab. The more protective mutations may have other, detrimental effects on the parasite that explain why they haven't spread as widely, Fidock says.

That competitive disadvantage could help slow the spread of the resistance outside of Asia. One reason that resistance spreads easily in Southeast Asia is that the region has relatively little malaria. A parasite that acquires a drug-resistant mutation doesn't have to compete with lots of other *Plasmodium* strains. In sub-Saharan Africa, where malaria is much more common, parasites have to outcompete other strains, both in people and in the mosquitoes that transmit the pathogens. So although researchers have found a number of K13 mutations in African parasites taken from patients, none has caused clinical drug resistance.

But as malaria control efforts continue to make progress, "there are many places in Africa where there is less and less transmission," says Carol Sibley of the University of Washington, Seattle, who is scientific director at the WorldWide Antimalarial Resistance Network. Latin America, too, has relatively little malaria and may be ripe for the spread of resistant strains, she notes.

The urgency of eliminating the resistant strains is greater than ever, Bozdech says. The *NEJM* paper in July showed that in spite of containment efforts, resistance is firmly established in Cambodia and had spread as far as Laos. If artemisinin resistance were to become widespread in Africa, "we'll be back where we were before World War II," he says. "All this progress will be erased." ■

PHOTOS: (TOP TO BOTTOM) PAULA BRONSTEIN/GETTY IMAGES; CDC

Published by AAAS

NUCLEAR POWER

The trouble with tritium

Japan searches for a way to cleanse contaminated water from the damaged Fukushima plant

By **Dennis Normile**, *in Fukushima Prefecture, Japan*

Hulking water tanks are popping up like gray mushrooms at the Fukushima Daiichi Nuclear Power Station. Nearly 1000 of the 10-meter-tall steel vessels hold more than 560,000 tons of radioactive water that has accumulated as ground water leaked into the wrecked reactor buildings and was pumped out. The plant operator, Tokyo Electric Power Co. (TEPCO), says it can now remove most of the radionuclides from the stored water—but not tritium. So, for the moment, the company dares not release the tainted water into the Pacific. "This contaminated water is the most urgent matter that we have to address," says plant manager Akira Ono.

A solution may be in sight. A U.S.-based firm says it has tweaked a catalytic process that, in principle, could concentrate all of Fukushima's tritium into just 5 cubic meters of water, at a cost of about $1 billion. The company is one of three that have won government grants to develop promising tritium-removal techniques. They have until March 2016 to demonstrate that their technologies can meet the Fukushima challenge.

Nearly 4 years after the 11 March 2011 earthquake and tsunami that led to meltdowns of three of Fukushima's six reactors, TEPCO is still in the early stages of a projected 40-year, $80 billion decommissioning task. Last month it removed the last of 1331 radioactive spent fuel rods stored in unit 4, clearing the way for its demolition. Next they need to find and retrieve damaged fuel. TEPCO determined that in one of the three stricken reactors, fuel debris likely melted its way through the reactor vessel and into the concrete containment building. But "we do not know exactly where those fuel rods are," Ono says. Intense radiation keeps human inspectors from venturing close to the epicenter of the meltdowns, but remotely operated devices may eventually be able to inspect and remove the damaged fuel.

In the meantime, crews must continuously circulate water through the buildings to cool the fuel and capture the 300 tons of ground water that leak in every day. After more than a year of struggling to purify the stored water with a balky system of filters and adsorbers, Ono says they are now reliably removing 62 radionuclides.

Removing the tritium is a steeper challenge. An isotope of hydrogen, tritium can replace a hydrogen atom in water and pass straight through the filters and adsorbers. Tritium's radiation risk is modest, as it gives off only low-energy beta particles. There would be "very limited human health concern from oceanic discharge," says James Seward, a radiation health specialist at the University of California, San Francisco.



This system strips nuclides from Fukushima's contaminated water but misses the tritium.

"This contaminated water is the most urgent matter that we have to address."

Akira Ono, Tokyo Electric Power Co.

The political risk is another matter. Japanese regulations allow nuclear power plants to discharge tritium in concentrations of up to 60,000 becquerels per liter into the environment. Virtually all nuclear power plants worldwide release tritium into the environment as part of normal operations. But they don't typically dispose of massive amounts all at once. Even a gradual release of Fukushima's tritiated water would likely be opposed by fishers who "worry that consumers won't be convinced" of the safety of seafood caught off Japan's northeast coast, says Jota Kanda, a biogeochemical oceanographer at Tokyo University of Marine Science and Technology.

Existing techniques for separating tritium from tritiated water are "very challenging, cost a lot, and take time," Kanda says. In December 2013, a Ministry of Economy, Trade and Industry advisory committee reported that no current technologies were suitable for use at Fukushima. In August, the ministry offered $8.4 million development grants to three groups with promising alternatives. One withdrew; the ministry is recruiting a replacement. A second group, FSUE Radioactive Waste Management Enterprise of Moscow, did not respond to *Science*'s requests for details. But Kurion Inc., of Irvine, California, was happy to talk up its project.

Kurion has modified a process called combined electrolysis catalytic exchange. It starts by splitting the contaminated water molecules into gaseous oxygen and a mixed gas of hydrogen and tritium. The hydrogen mixture flows into the bottom of a column filled with a platinum catalyst while water trickles in from the top. The catalyst triggers a swap in which a hydrogen atom in the water trades places with a tritium atom in the gas. Clean hydrogen flows out the top of the column; water containing a higher concentration of tritium collects at the bottom.

The process is similar to one used to cleanse contaminating tritium from the heavy water used in certain nuclear reactors. But Gaëtan Bonhomme, Kurion's chief technology officer, says his firm has optimized the column's design so that one trip through squeezes the tritium into 25% of the original volume of water entering the electrolyzer. Repeatedly recirculating the water further concentrates the tritium.

After the development program wraps up in early 2016, the ministry and TEPCO will decide on their next step. Ultimately, says Masao Matsuyama, director of the Hydrogen Isotope Research Center at the University of Toyama, how to handle the tritium "will be a decision for TEPCO and the politicians." ■

PHOTO: EPA/SHIZUO KAMBAYASHI/POOL PHOTO

Downloaded from www.sciencemag.org on December 14, 2014

Published by AAAS



Ancient asteroid impacts may have kick-started life on Earth

ORIGIN OF LIFE

From hell on Earth, life's building blocks

Simulated meteorite impact forges RNA components

By **Sid Perkins**

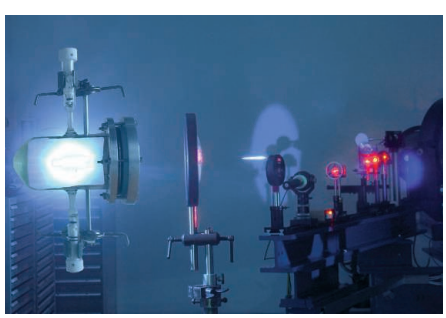
A high-powered laser in the Czech Republic has now provided provocative evidence that the hellish conditions produced when an asteroid or comet slams into Earth could have created some key building blocks of life on Earth. In a lab experiment intended to duplicate the high temperatures and pressures of such an impact, researchers transformed a solution containing a simple precursor into adenine, guanine, cytosine, and uracil—the information-bearing nucleobases in RNA, which many believe to have been the first genetic molecule to encode.

"This is, I believe, the first time that all four nucleobases have been made in one set of reaction conditions," says Steven Benner, an astrobiologist at the Foundation for Applied Molecular Evolution in Gainesville, Florida.

Researchers have long sought to identify ways that the nucleobases that make up RNA or DNA could be created from simpler substances. That, could help scientists ascertain where and how life might have originated, says Svatopluk Civiš, a physical chemist at the J. Heyrovský Institute of Physical Chemistry in Prague.

In recent years, Civiš adds, researchers proposed that a substance called formamide was a possible source of such genetic building blocks. This minimalist chemical, which forms when hydrogen cyanide reacts with water, would have been abundant on early Earth and has the major elements needed for prebiotic chemicals—namely, hydrogen, nitrogen, carbon, and oxygen. Indeed, some teams have already produced individual nucleobases in lab experiments that relied on catalysts to drive chemical reactions between formamide and other ingredients. Other teams have made nucleobases from different grab bags of simple chemicals.

In the new research, Civiš and his colleagues fired their institute's 1-kilojoule laser into a formamide-bearing solution that also included clay. In that mixture, intended to represent a chemical-rich pool on ancient Earth's surface, the one-third-of-a-nanosecond-long pulses generated intense pressure, temperature spikes exceeding 4200°C, and a cascade of radiation including ultraviolet and x-ray wavelengths—just the sort of conditions expected when an object such as a comet or asteroid strikes the ground. These extreme conditions sparked reactions that, besides producing substances such as hydrogen cyanide, carbon monoxide, ammonia, and methanol, also created the four RNA nucleobases, the researchers report this week in the *Proceedings of the National Academy of Sciences*.



A laser blasting a solution of the chemical formamide (large white glow) generated RNA's nucleobases.

In addition to being a nice bit of lab work, Benner says, "this paper has real prebiotic relevance since it uses as a starting material a compound that was undoubtedly present in sizable amounts on early Earth."

Previous studies showed that some classes of meteorites already contain nucleobases such as adenine and guanine. But the new results suggest that besides merely delivering nucleobases, celestial bodies could have also created them when they struck the planet, says Raffaele Saladino, an organic chemist at Tuscia University in Viterbo, Italy. He and his colleagues have studied formamide under more benign lab conditions, but he says the Prague experiments "are the first time that formamide has been studied in the context of a meteorite impact."

Such impacts were common in Earth's early history. During a period aptly dubbed the Late Heavy Bombardment, which began about 4 billion years ago and lasted some 150 million years, large objects pummeled our planet and moon as well as Mercury, Venus, and Mars. Although many researchers have suggested that such impacts and their resulting effects on climate would have effectively sterilized Earth's surface, erasing any life that may have already started, the new study hints that this period seeded our planet with the raw ingredients necessary for life to develop. "This paper nicely correlates the Late Heavy Bombardment and the energy it delivered to Earth about 4 billion years ago with the formation of RNA and DNA nucleobases from formamide," Benner says. ■

Sid Perkins is a writer based in Crossville, Tennessee.

CREDITS: (ILLUSTRATION) DANA BERRY/NATIONAL GEOGRAPHIC CREATIVE; (PHOTO) FERUS *ET AL.*, *PNAS* (2014)

Downloaded from www.sciencemag.org on December 14, 2014

Published by AAAS

EUROPE

E.U. to let wary members ban genetically modified crops

Revived proposal allows countries to reject varieties even if scientific assessments say they're safe

By **Tania Rabesandratana**, *in Brussels*

Some call it a defeat for science in the European Union—but others argue it's just a pragmatic way out of long-standing, irreconcilable differences. Starting in 2015, individual E.U. countries may be able to ban genetically modified (GM) plants—even if the European Union's own regulators have declared them safe for cultivation. Critics say the move undermines the entire concept of science-based risk assessments. But many member states welcome the plan, because it will allow them to follow the will of their citizens, who widely reject GM foods.

The proposal to "renationalize" decisions on GM crops, stalled for years, took a decisive step forward on 3 December, when negotiators representing the European Parliament and national governments agreed on a joint version of the bill. If the whole Parliament and member states formally endorse it in the coming weeks, it may come into force in the spring.

At the moment, GM crops are approved at the European level. The European Food Safety Authority (EFSA) in Parma, Italy, assesses applications from biotech companies for risks to public health and the environment. Based on EFSA's advice, the European Commission then decides if a crop should be allowed for cultivation. Member states have to sign off on the commission's decisions through a regulatory committee, made up of experts from each government.

Wide rifts between E.U. countries have slowed that process to a crawl. Only one GM crop, Monsanto's MON810, a pest-resistant maize, is now grown on E.U. fields. Many others have become stuck. A pest-resistant maize called 1507, for instance, for which DuPont Pioneer first sought authorization in 2001, was greenlighted by EFSA, but the member states' committee failed to reach a consensus and the commission sat on the file for several years. When it referred the matter to the higher level—the Council of Ministers—two-thirds of member states rejected the crop. But because each country has a different number of votes in the council's ballot system, opponents did not reach the "qualified majority" needed to turn down 1507, meaning that the commission is now allowed to authorize it—which it still hasn't done. Several countries have banned GM crops (see map) in defiance of E.U. rules—even MON810 is grown in only five countries—though some of the bans have come under legal fire.

0.15
GM maize area in the E.U. in 2013 (million hectares)

35.6
GM maize area in the U.S. in 2013 (million hectares)

The new plan gives GM opponents such as France and Austria the option to bypass the commission's decisions and keep a newly developed crop off their fields without having to prove EFSA wrong with fresh scientific evidence or running into legal headaches. "This is not about opposing science and politics … or about opposing reason and emotion," says Patrice Audibert, an aide to the Parliament's main negotiator Frédérique Ries, but rather about unlocking an untenable situation and enabling governments to respond to European consumers' mistrust of GM foods. At the same time, GM-friendly nations such as Spain and the United Kingdom certainly hope that having this option will make GM opponents more likely to sign off on pan-European authorizations.

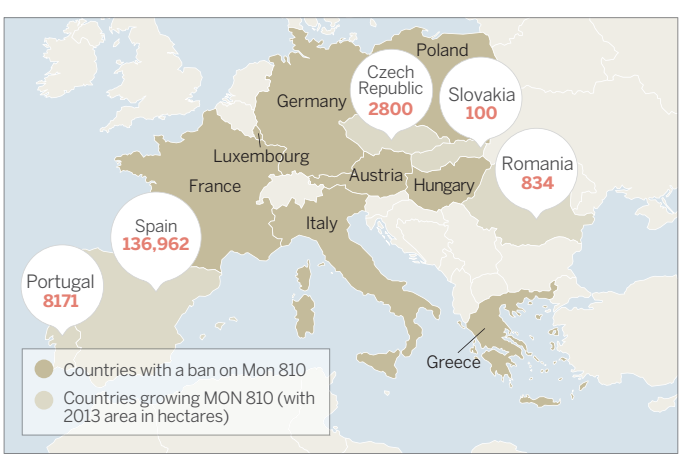
The bill has come under fire from many sides, however. Biotech companies say the basic idea behind it—an agreement to disagree at the expense of market harmony—sends a terrible message to industry and could hamper the introduction of other innovations. "Rejecting modern technologies on nonscientific grounds sets a dangerous precedent … and sends a negative signal for innovative industries worldwide considering whether or not to invest and operate in Europe," said Beat Späth, director for agricultural biotechnology at the industry association EuropaBio here, in a statement. Meanwhile, green groups say the bill does not give sufficient legal backing for national bans. It mentions maintaining local biodiversity and ecosystems as reasons that member states could invoke to ban the crops, but Marco Contiero, agriculture policy director at Greenpeace EU here, says the wording is not enough to withstand legal scrutiny if seed producers challenge a ban in court.

Scientists on both sides of the GM issue criticize the proposal as well. Angelika Hilbeck, a biologist at the Swiss Federal Institute of Technology in Zürich and a GM critic, says that the plan is a cop-out and "a political solution to the fact that there is no scientific consensus on safety and that we are stuck in gridlock." Hilbeck calls the plan a "face-saving attempt [to rescue] EFSA's authority and credibility" that fails to address scientific dissent.

But Stefan Jansson, a professor of plant cell and molecular biology at Umeå University's Plant Science Centre in Sweden, says that EFSA's risk assessments are thorough and that there is no real scientific debate about the safety of GM crops. From a scientific point of view, the European Union's plan "is a step backwards," Jansson says. "But if it helps to break the deadlock, maybe it is progress anyway. Time will tell." ■

A divided union

MON810 maize is the only GM crop grown in the E.U.; the amounts are tiny compared to the U.S., and eight countries have banned the variety.



DATA: GMO COMPASS; GMO-FREE EUROPE

Downloaded from www.sciencemag.org on December 14, 2014

Published by AAAS



Downloaded from www.sciencemag.org on December 14, 2014

BACK FROM THE DEAD

The once-moribund idea that volcanism helped kill off the dinosaurs gains new credibility

By **Richard Stone**, *in Maharashtra state, India*

Gerta Keller's latest findings have pulled her from the fringe back to the frontier.

PHOTO: PETER MURPHY

Perched on a bluff on the Deccan Plateau here in southwestern India, the ruins of Sinhagad Fort command a sweeping view of sienna-hued volcanic ridges draped with haze from the nearby metropolis of Pune. Turn back the clock nearly 66 million years, and this vantage point was hell on Earth. Then, India was an island straddling the Réunion hot spot: a magma fountain gushing from deep in the earth that, for some 750,000 years, flooded an area the size of Mongolia with lava in one of the largest episodes of volcanism in Earth history.

Today this rugged terrain is ground zero for a long-running and bitter scientific debate, centered on the cataclysmic extinctions that took place at roughly the same time. One camp argues that Deccan volcanism spewed enough carbon dioxide to send temperatures soaring globally while sulfur fumes turned the oceans to acid, with lethal results for three-quarters of Earth's life forms, including all nonavian dinosaurs. The other camp, in the ascendancy for decades, blames the die-off on a single killer: the impact of an asteroid.

Online this week in *Science* (http://scim.ag/BSchoene), volcanism gets a big boost, and so does one of the most contentious figures in the debate, Princeton University paleontologist Gerta Keller. In the paper, she and her colleagues report precise dates for the main phase of Deccan volcanism. Countering claims that the lava outpouring took place too early to have played any role in the extinctions, they show that it straddles the Cretaceous-Paleogene boundary (KPB): the geological moment when many creatures from the age of dinosaurs vanish from the fossil record. "It's an important paper," says Paul Renne of the Berkeley Geochronology Center in California. "It shows that most of the action happened over a limited time range" and that older ages ascribed to the Deccan flows are wrong.

Published by AAAS

The new findings mark the latest step toward redemption for the volcanism hypothesis. The idea was sidelined in 1991, when remnants of a 300-kilometer-wide crater called Chicxulub were confirmed off the Yucatán Peninsula. This is the scar left by a 10-kilometer-wide asteroid or comet that collided with Earth a mere 32,000 years before the Cretaceous ended. (Renne and others nailed the impact's timing in a report in *Science* last year.) Since then, however, growing evidence has hinted that the asteroid might not have been the only killer at work. Of a dozen enormous lava flows known as flood basalts in the geological record, four immediately preceded mass extinctions. And many paleontologists argue that well before the Chicxulub impact brought the curtain down on the Cretaceous, a host of species had already winked out or were in sharp decline. "The pendulum has certainly swung" toward a role for volcanism, Renne says.

More and more scientists now embrace the notion of a one-two punch at the KPB: Deccan volcanism crippled many species and extinguished some before the Chicxulub bolide came along to deliver a knockout blow. "The new paper in no way diminishes the role of the impact," says Shen Shuzhong, a paleontologist at the Nanjing Institute of Geology and Palaeontology in China. But a link between Deccan volcanism and the mass extinction, he says, is now "hard to deny."

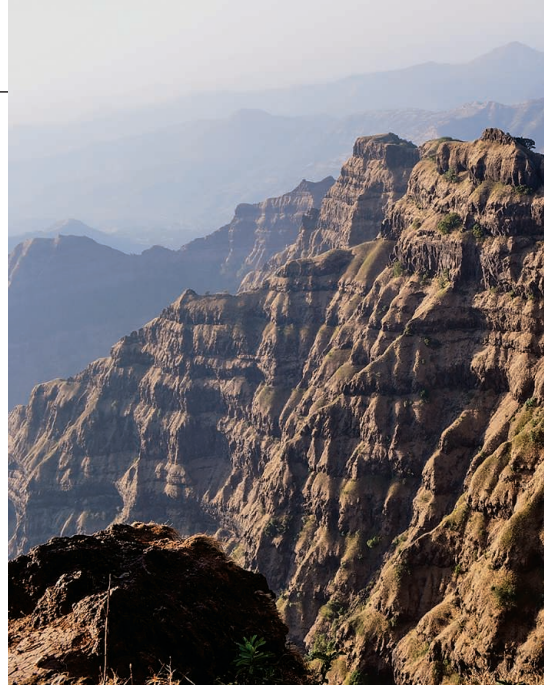
Hard, maybe, but not impossible. Key figures in the impact camp refuse to countenance any role for volcanism—and are unswayed by the new findings. "The crux of the matter is this: Is Deccan a factor in the extinctions? I say it's not," says Alan Hildebrand, a geologist at the University of Calgary in Canada. "What I see are people desperately trying to find something that doesn't fit the impact hypothesis." Impeding any armistice is the divisive figure of Keller herself.

"I MAY BE A BIT STUBBORN," Keller says. "But there is a silent majority who knows that I am right." Now 69, Keller grew up on a dairy farm in Switzerland with 11 brothers and sisters. In 1964, desperately poor and believing she had no future at home, the teenager took off backpacking for several months and ended up in Australia. In Sydney, she says, she was shot by a bank robber attempting to carjack her Volkswagen. As she lay near death, a priest twice urged her to confess her sins. "I said, 'No, I have nothing to confess.' That made me mad, and probably saved my life." A few years later, she left for the United States, and in 1978 she earned a Ph.D. in paleontology and geology from Stanford University.

Keller got her start at a heady time for the geosciences. In 1980, the father-son team of Luis and Walter Alvarez and two colleagues published a landmark paper in *Science* proposing an "extraterrestrial cause" for the KPB extinctions. Their smoking gun was high levels of iridium turning up in KPB sediments. Because the element is depleted in Earth's crust but abundant in meteorites, they argued that it had to have originated in an asteroid that collided with Earth and flung iridium-laden debris across the planet.

The provocative idea captured the public's imagination but got a cool reception from scientists who favored a gradual extinction trigger, such as protracted volcanism. Paleontologists, for instance, argued that the fossil record hinted at a prolonged die-off, not a sudden death blow. The dispute quickly became personal. Luis Alvarez, who had won the Nobel Prize in physics in 1968 for his discovery of resonance states in subatomic particles, famously ridiculed paleontologists as "stamp collectors." One proponent of a volcanic explanation for the KPB extinctions, paleontologist Dewey McLean of the Virginia Polytechnic Institute and State University in Blacksburg, says his encounters with the elder Alvarez in the early 1980s were so venomous that they wrecked his career and his health. "Luis was not a gentle person," Renne says. "A lot of people with opposing views were pushed around."

After moving to Princeton in 1984, Keller ventured into the maelstrom, believing that fossil marine life, among other indicators, didn't square with the impact theory. "The data showed that the mass extinction could not be instantaneous," she says. Her first run-in with the impact camp came in 1988, at a meeting on global catastrophes in Snowbird, Utah. When she started to deliver her talk, she says, "I didn't even get past my introduction and I was already being shouted down." Anyone who had the temerity to question the impact theory, she says, "was simply told, 'You don't know what you're doing. This is all stupid, it's ridiculous, blah, blah, blah.' "

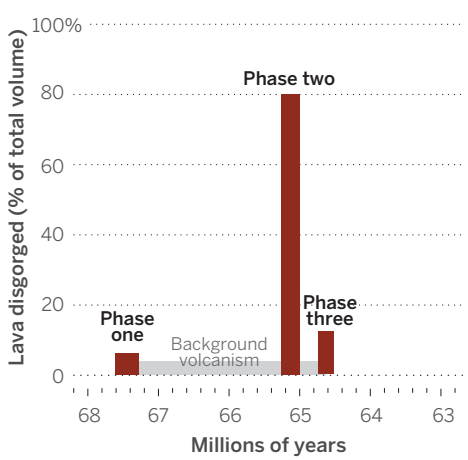


Keller herself was developing a reputation for obduracy. As a graduate student in the 1980s, Hildebrand, energized by the impact hypothesis, set off in search of a crater. "Gerta made a big point of telling me that I was wrong—that I was wasting my time," he says. Throughout that decade, Hildebrand and others turned up rock ejected by the impact in Haiti and elsewhere, but the crater eluded them. In fact, it had already come to light, although almost no one had noticed. In 1978, geophysicists surveying for an oil company had discerned a vast, buried scar centered in the Gulf of Mexico near the Yucatán village of Chicxulub.

Sea of fire

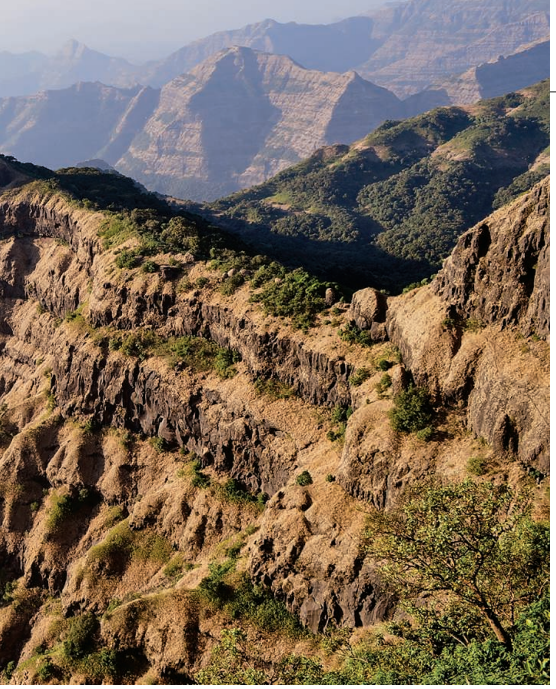
The lava flows of the Deccan Traps span an area the size of Mongolia. New findings show that most of the lava was disgorged over a 750,000-year period precisely when many denizens of the age of dinosaurs vanish from the fossil record.

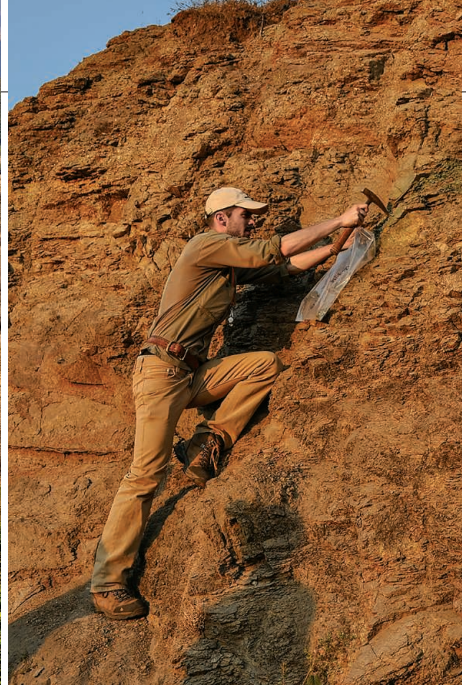




CREDITS: (PHOTO) GERTA KELLER; (DATA) GERTA KELLER

Published by AAAS





Team member Mike Eddy samples a layer that could contain zircon-rich ash from the awe-inspiring Deccan Traps near Pune.

PHOTO: GERTA KELLER

Hildebrand got wind of the find in 1990 and contacted one of the scientists; together they obtained core samples from the oil company and found shocked quartz and glass spherules that could only have been forged in the cauldron of a bolide impact or a nuclear explosion. Teams fanned out across Yucatán and found outcrops left by the geological calamity, including rhythmically layered sandstone, a couple of meters thick, just below the iridium-rich stratum. Several researchers suggested that the layers were deposited by a gargantuan tsunami unleashed when the bolide slammed into the ocean.

When the tsunami evidence looked overwhelming, Keller was ready to move on. "I thought I'd had it by then," she recalls. But she says that Charles Officer, a geologist now retired from Dartmouth College and an implacable impact critic, badgered her for days to make a field trip to Yucatán to inspect the KPB outcrops for herself. She relented, but her adversaries refused to give her the coordinates. That only stiffened Keller's resolve.

Mexican researchers helped her locate the crater's KPB layers, and several dozen field trips to Yucatán over the next decade generated extraordinary claims. She and her colleagues described finding fossils of tiny aquatic life forms and their burrows in the sandstone, suggesting that the sedimentary rock attributed to the tsunami had actually formed over thousands of years. Later, she also claimed that the Chicxulub impact occurred 300,000 years before the KPB and that its effects would have been too short-lived to trigger the die-off. Most startling, she asserted that multiple, undiscovered impacts contributed to the KPB iridium—but that much of it is of earthly origin.

"She's been fighting the Chicxulub crater" ever since it was found, Hildebrand says. Jan Smit, a geologist at VU University Amsterdam, scoffs at the claims of life forms and burrows in the tsunami sediments: Keller's group mistook dolomite crystals for fossils, he asserts, while the burrows were made "after the tsunami waves subsided." And even Keller's backers shake their heads at the multiple impact theory. To Smit and others in the impact camp, Keller had strayed into the fringe.

BELOW THE RAMPARTS of Sinhagad Fort, Syed Khadri briskly strikes a volcanic outcrop with a rock hammer. Trapped inside, he hopes, are clues to what killed the dinosaurs. "There are no fossils in this basalt, of course," says the geologist, from Sant Gadge Baba Amravati University in India. "We're looking for zircons."

A few months earlier, Keller, Princeton geochronologist Blair Schoene, and a few others had joined Khadri on a zircon hunt. When zircons crystallize, they trap uranium atoms, which turn the tiny crystals into precise timepieces that run for billions of years. The radioactive element decays into lead at a known half-life, so the ratio of the two elements' isotopes reveals how long ago the crystal formed. Zircons are scarce in basalt but common in the silica-rich ash of explosive volcanic eruptions. Schoene's team found such ash—overlooked by past geochronologists, he says—embedded between layers of flood basalt. The zircons he and his colleagues culled have allowed them to conclude that the main eruptions began about 250,000 years before the KPB and lasted roughly 750,000 years.

Smit contends that the date Schoene and company arrived at for the onset of the main phase—66.29 million years ago—could be spurious. It may be "from a much younger vein [of lava], crossing the old lava flows like a sill," he says. That's "very unlikely," Schoene says: Sills are seldom seen in the Deccan Traps, and the layers his group sampled seem to have been deposited in chronological order. He adds that the dates "perfectly" match earlier paleomagnetic work led by Vincent Courtillot, a professor emeritus of geophysics at Université Paris Diderot and the Institute of Earth Physics of Paris. "Of course, more datable samples from the base of the Deccan are the best way to answer Jan's skepticism," he says.

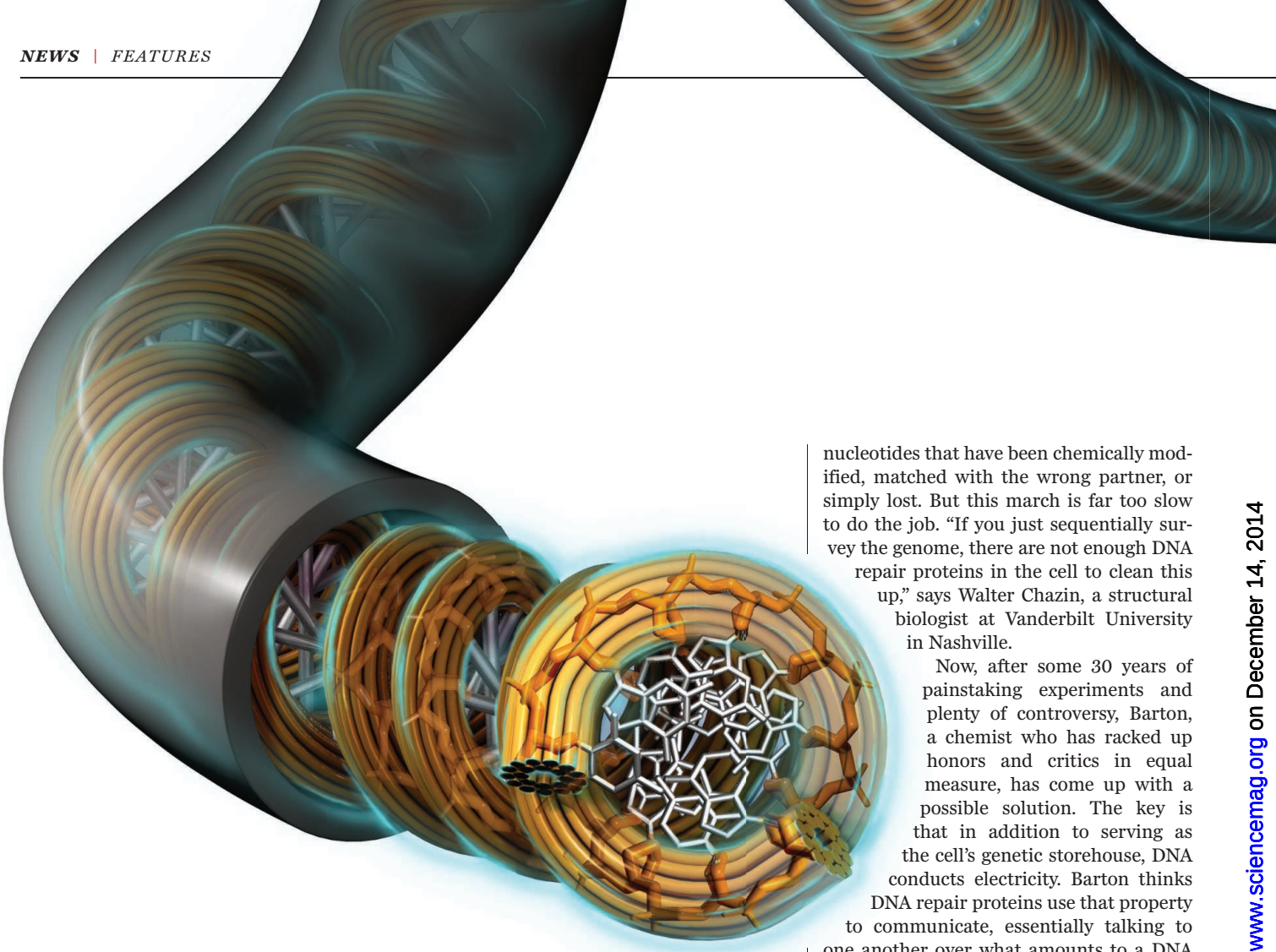
Smit and Hildebrand dispute another central tenet of the Deccan hypothesis: that many species were in trouble before the Chicxulub impact. There's "lots of arm waving," Smit says, but "no evidence whatsoever." He doubts that the Deccan eruptions "had more influence on biota than just a few minor temperature changes." In contrast, he and Hildebrand say, the impact's legacy in the fossil record was instant and dramatic: Marine species vanish abruptly, and ferns, the ultimate survivor, dominate the land record right after the catastrophe.

The Princeton team concedes that their latest result doesn't clinch the case. For example, it does not reveal whether the eruptions took place as a steady trickle or as brief, massive outpourings, which would have had a much more profound impact on the environment. "We really need to get at the tempo," Renne agrees.

Courtillot, a longtime champion of Deccan volcanism, says paleomagnetic data and other evidence suggest that massive lava pulses—some exceeding 10,000 cubic kilometers, 100 times larger than any historic eruption—were released in a flash: possibly less than a decade. "The gas output of such flows was comparable to that of the impact, and there were many such flows," he says. To test that idea, Schoene and Keller are heading back to the Deccan Traps next month in search of more zircons, which they hope will yield finer detail about the duration and size of the lava pulses. "If we can nail down the flows that initiated the extinctions," Keller says, "then we really are in business."

The next step would be to show exactly how those eruptions might have wiped out species. "We need to explain how you get the global mass extinction from volcanism," she says. Her team is now looking for evidence of pulses of ocean acidification accompanying Deccan eruptions in the 50,000 years before the KPB. Already, though, many of her peers say that the new findings have pulled Keller from the fringe back to the frontier. At least, she says, with a laugh, "I don't expect hostile receptions anymore." ■

Published by AAAS



LIVE WIRE

Do cells use electricity to repair DNA? Jacqueline Barton aims to find out

By **Robert F. Service**

"It's a crazy idea, right?" says Jacqueline Barton. Sitting composedly in her bookshelf-lined office at the California Institute of Technology (Caltech) in Pasadena, she looks the picture of the establishment scientist. But for decades, she has fought a battle with many of her biochemist colleagues over the properties of DNA and, more recently, her unorthodox proposal about how the body repairs damage to this vital molecule.

Up to a million times a day, photons in sunlight; mutagens in food, water, and air; and other assaults damage the DNA molecules coiled inside the nucleus of each cell. A broad family of protein sentinels perpetually scans our genomes and repairs any damage they find before it leads to cancer or other diseases. But just how these repair proteins manage to find the damage amid the genome's 3 billion pairs of nucleotides remains a puzzle. The repair proteins work by hopping onto the DNA and marching along the double helix scanning for nucleotides that have been chemically modified, matched with the wrong partner, or simply lost. But this march is far too slow to do the job. "If you just sequentially survey the genome, there are not enough DNA repair proteins in the cell to clean this up," says Walter Chazin, a structural biologist at Vanderbilt University in Nashville.

Now, after some 30 years of painstaking experiments and plenty of controversy, Barton, a chemist who has racked up honors and critics in equal measure, has come up with a possible solution. The key is that in addition to serving as the cell's genetic storehouse, DNA conducts electricity. Barton thinks DNA repair proteins use that property to communicate, essentially talking to one another over what amounts to a DNA telephone line to dramatically speed their search for damage. "This communication is the first step in repair," she says, clicking through slide after slide in a presentation that outlines dozens of studies her group has conducted over the years.

As if that proposal weren't enough to provoke the skeptics, Barton is now going further. Recent evidence suggests to her that DNA telephony not only speeds DNA repair but may also play a role in most processes in which proteins manipulate DNA—even the all-important tasks of reading and copying genes. "This is all incredibly coordinated," Barton says. "They have to talk to each other. You need that in a cell, and in a genome. … The telephone game makes everybody better."

"It's an incredible idea," says Chazin, who originally counted himself among the skeptics, but is now actively researching it in his own lab. Chazin and others note that many biologists still balk at the notion that proteins would inject electrons into DNA, potentially damaging the very molecule they are meant to safeguard. But he adds, "The trajectory that DNA charge transport plays a biological role is on the rise."

ILLUSTRATION: V. ALTOUNIAN/*SCIENCE*

Downloaded from www.sciencemag.org on December 14, 2014

Published by AAAS



In pristine DNA, parts of the nucleotide core conduct electricity like a molecular wire. But when DNA gets damaged, its conductivity drops—a property cells may harness to home in on problems.

THE NOTION that DNA could conduct electricity actually dates back to shortly after Watson and Crick solved the structure of the molecule in 1953. Chemists noted that the double helix was only part of what made DNA interesting. The sugar and phosphate groups that make up DNA's "backbone" actually more closely resemble an exoskeleton, spiraling around a core of nucleotide bases. These bases pair up, As binding to Ts and Gs to Cs. Each base contains a ring of atoms, known as an aromatic group, with doughnut-shaped electron clouds above and below each ring. The aromatic groups and their electron clouds overlap, creating what looks like a continuous path for electrons—a molecular wire, if you will. But in the 1950s and 1960s, when chemists looked for signs that DNA conducts electrons, most found nothing.

Barton knew nothing of this history in 1983, when she moved from Hunter College to Columbia University as a young professor. At the time, she was studying the transfer of charges between metal compounds capable of binding DNA at specific sites. In 1985, one of her postdocs, Vijay Kumar (now at the University of Connecticut, Storrs), added some DNA and saw conductivity shoot up, suggesting that, like a wire, the DNA was helping charges move between distant metals. Looking back through the literature, Barton and Kumar spotted the earlier studies. In most of them, researchers were testing DNA that had been isolated, dried, and frozen for easier analysis. "People weren't looking at DNA under physiological conditions," Barton says. So she and her colleagues measured DNA's conductivity at room temperature and in solution and saw unmistakable signs that DNA was in fact ferrying charges over long distances.

Researchers studying another class of biological molecules, proteins, were finding nothing of the kind. Proteins typically transfer electrical charges only very short distances—at most a few nanometers—through a process known as quantum tunneling. But if Barton's studies were correct, DNA was ferrying charges dozens of nanometers, as much as 100 base pairs down the line, far beyond tunneling's reach. "People at first were quite skeptical because the distances [seen by Barton] were so long," says Harry Gray, a chemist at Caltech, who at that time was studying electron transfer in proteins and was cautious about the new DNA results. Among other things, the skeptics contended that metals inserted into DNA for her studies were causing experimental artifacts. "It was very painful," Barton says of the early criticisms. "But it forced us to do better experiments."

They did, and steadily amassed evidence that DNA really could behave like a wire. In one 2008 study, for example, Barton teamed up with Columbia chemist Colin Nuckolls to build transistors in which DNA was wired between two carbon nanotubes, which themselves were connected to separate electrodes. They found that when all the DNA bases were paired up normally, charges readily flowed through the DNA wire. Yet when there were mismatches in the pairing (say, a C paired with another C instead of a G), DNA's conductivity dropped sharply. Barton's team has since shown that changes in conductivity can reveal multiple kinds of DNA damage, even when it involves just a single base. Even subtle changes apparently disrupt the perfect stacking of nucleotide bases required for charges to flow swiftly. "What we learned was that DNA's abil-



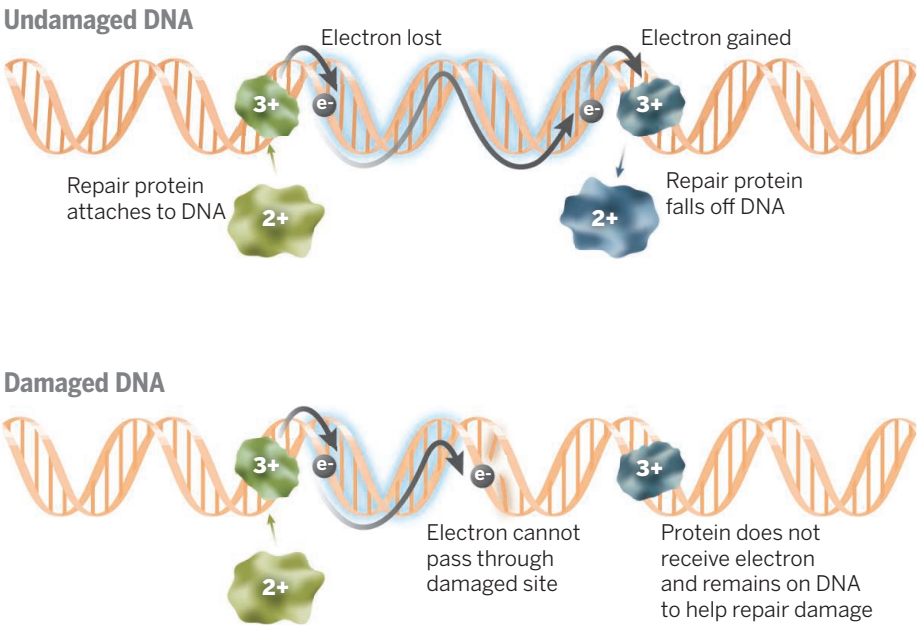
Barton has grown accustomed to critics uncomfortable with her ideas.

PHOTO: DAVID WALTER BANKS

Published by AAAS

How a "telephone game" might pinpoint DNA damage

Repair protein (green) latches on to DNA and sends in an electron. If the DNA is undamaged (*top*), the electron travels until another repair protein (blue) picks it up and falls off. If the DNA is damaged (*bottom*), the second protein doesn't pick up the electron and keeps scanning for damage.



ity to conduct charges depends exquisitely on the stacking of the bases," Barton says. These and other studies have largely won over her early critics. "I'm on board," Gray now says. "I think the community has pretty much come around to agreeing that there is charge transport over long distances."

BUT IS THAT MORE THAN A CURIOSITY? Researchers can't clamp alligator clips onto DNA inside the nucleus to find out whether it is conducting electricity in living cells. So Barton's group had to look for indirect evidence. One starting point was a 1992 *Science* paper by John Tainer, a structural biologist at the Scripps Research Institute in San Diego, California, and colleagues. Tainer's team was the first to work out the crystal structure of a DNA repair protein known as endonuclease III. EndoIII is a member of a broad family of repair proteins that carry out a process known as base excision repair (BER), snipping out damaged nucleotide letters so that other repair proteins can come in and insert a pristine copy.

When Tainer's team looked at EndoIII's structure, they were puzzled. In addition to the usual twisting chain of amino acids, the protein also harbored a cluster of four iron and four sulfur atoms (Fe_4S_4). In many enzymes, iron-sulfur clusters play a key catalytic role because they are adept at either grabbing an electron from their neighbors or giving one up, chemical processes known as reduction and oxidation (redox). In other proteins the metal clusters help the proteins fold into the proper shape. But when Tainer looked at the structure of EndoIII, the Fe_4S_4 cluster didn't seem to be in the right place to help the protein either fold or catalyze a reaction. Even more puzzling was that, because it is a DNA repair enzyme, EndoIII regularly comes into contact with DNA. That meant evolution was putting a potentially electron-conducting metal group right next to its genetic vault. "That's exactly what you wouldn't want," Tainer says. So just what is the iron-sulfur cluster doing there?

> ***"Jackie Barton did the right experiments. ... [She] opened the door for big changes in our understanding."***
>
> **John Tainer**, Scripps Research Institute

In hopes of finding out whether the metal group's electronic properties were playing some role, Tainer and his colleagues isolated samples of EndoIII and exposed them to other compounds, measuring how likely the protein is to give up or grab electrons, a property known as its redox potential. They found that only strong redox compounds had any effect—compounds that aren't present in living cells. Tainer's team concluded that EndoIII's Fe_4S_4 group wasn't exchanging electrons inside the cell, but probably played a structural role. To Tainer that looked like the end of the story.

But Barton wasn't satisfied. She wondered if the protein's redox potential changed when it latched on to DNA. To find out, she and her colleagues tethered copies of DNA to electrodes in solution and added EndoIII. They applied different voltages and watched to see whether the protein's Fe_4S_4 group either gave up or snagged an electron. When EndoIII was bound to DNA, the protein's redox potential plummeted right into the sweet spot for biological activity.

Tainer was convinced. "In my view, Jackie Barton did the right experiments," he says. She "opened the door for big changes in our understanding of protein-DNA transactions and regulation."

Barton's team settled on a radical interpretation of the results: Repair proteins might be using DNA to talk to one another. Popping up from her chair, she walks over to the monitor to point out what's going on. As she pictures it, when a repair protein binds to DNA, it changes its configuration, making the Fe_4S_4 group more apt to give up an electron. When it does, the protein's charge increases from 2+ to 3+, holding it more tightly to the DNA. The ejected electron, meanwhile, zips down the DNA wire. If that electron runs into a damaged bit of DNA, it can't proceed. But if not, it keeps moving until it encounters another DNA repair protein down the line. "This second protein grabs the electron and goes from [a charge of] 3+ to 2+, and falls off the DNA," Barton says. In short, a mini telephone game between the two proteins signals the protein on the receiving end that all is clear upstream and that it's okay to move on to look for damage elsewhere. "If these guys can talk to each other, it shows there is no problem with the DNA in between. It makes the search more efficient," Barton adds.

In that scenario, Barton and her colleagues reasoned, repair proteins would tend to cling to damaged DNA, because they would not receive the electrons that could free them. The team found just that. In a pair of papers in 2009 and 2012 in the *Proceedings of the National Academy of Sciences*, they reported using an atomic force microscope to count the distribution of base excision repair proteins on DNA strands that either were undamaged or had one damaged base. They found that, on average, damaged strands had many more repair proteins attached to them than flaw-

CREDITS: (ILLUSTRATION) C. SMITH/*SCIENCE*; (SOURCE) ADAPTED FROM JACKIE BARTON

Published by AAAS

less strands did. What was key, Barton says, is that the BER proteins congregated on the damaged strands even though the types of damage weren't ones they were designed to fix. In a separate study, the team mutated the repair proteins and showed that those with mutations that made them less adept at transferring electrons were not as efficient at finding damaged DNA.

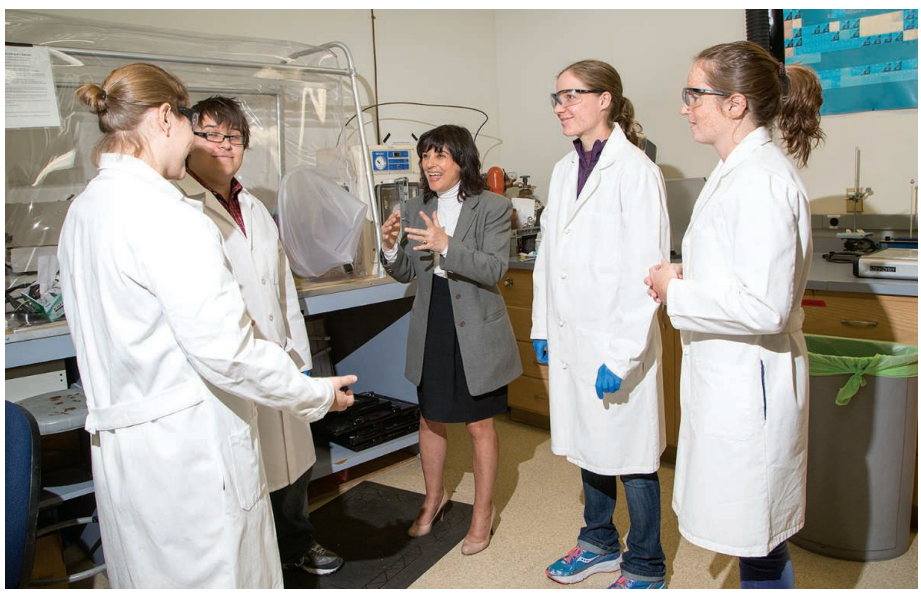
Barton's team also experimented with removing or altering the genes for repair proteins in bacteria to see whether the repair proteins collaborate the way her team envisions. When they knocked out the gene for EndoIII, for example, they found that the repair efficiency of a separate protein called MutY declined. When they reinserted the EndoIII gene, MutY regained its mutation-fixing abilities. They repeated the experiment with a gene for a modified EndoIII that could transport charge normally but could not cut damaged DNA. MutY's efficiency still improved. "It shows they communicate," Barton says. "Endo is helping MutY find its lesions even though they are doing different jobs." Barton's team has since reported similar results with other groups of repair proteins. Earlier this year, for example, they reported in the *Journal of the American Chemical Society* that bacteria prone to DNA damage can grow and replicate only when charge-transport proteins capable of playing the telephone game are restored.

"Taken together it's pretty compelling," Tainer says. Caltech's Gray agrees. "The evidence is piling up that there is real importance biologically," he says. But Gray, Chazin, Tainer—and even Barton—acknowledge that all of these tests on modified organisms and in vitro still leave room for doubt about whether proteins use DNA to play the telephone game in real life. "I would say this has not been proven," Chazin says.

Critics go further. "I think a lot of people don't buy this electron transfer stuff," says one U.S. biochemist specializing in DNA repair who was unwilling to give his name for fear of seeming combative. One problem, he says, is that the amino acid chains of protein are typically insulators, making it hard to see how DNA repair proteins could readily inject electrons into DNA or receive them. He notes that in their experiments, Barton's group often wedges an electron-donating group into the DNA helix or places the helix in contact with an electrode, steps that could artificially boost charge transfer. "I'm a little worried she gets ahead of the facts," the anonymous critic says. Other outsiders contacted by *Science*, several of them biologists, expressed similar concerns but were also unwilling to voice their doubts publicly. One biologist said, "I think people are reluctant to comment on this because they are either overburdened or do not like to cast stones" at a respected researcher, who they believe is working outside her field of expertise.

In the face of such skepticism, "it took a lot of courage and insight for Jackie to push this area," Tainer says. "Scientists tend to be conservative," he adds. "I think a lot of people don't want to find out they were missing a big piece of the picture." For her part, Barton says dealing with persistent criticism has been painful at times, but she insists she's only following the evidence. "I like data," she says with a broad grin.

That single-mindedness and her positive attitude have won her a legion of fans, particularly among her students. "When I talked to Jackie [for the first time], she had a level of excitement that was very intense. It automatically attracts you," says Natalie Muren, one of Barton's postdoctoral assistants.



Barton, seen with her students, argues that not all proteins that work with DNA play the "telephone game." But those that do speed the processes that find and repair damage.

PHOTO: DAVID WALTER BANKS

Barton's work on DNA's role as a molecular wire and her mentorship have also won accolades. In June she was named the 2015 Priestley medalist, the highest honor given by the American Chemical Society. And in 2010 she won the National Medal of Science, which she describes as "a lump-in-your-throat experience." (Barton's husband, Peter Dervan, also a chemist at Caltech, won the same award in 2006, making Barton and Dervan the only husband-and-wife pair of laureates.)

The telephone game played by repair proteins might be only the beginning of the story. Spurred by her work, biochemists are looking elsewhere for iron-sulfur clusters, which have begun turning up in a wide variety of enzymes that work on DNA—including the polymerases that copy it, helicases that unwind it so it can be copied, and primases that help polymerases start their copying work. "Any time DNA needs to be processed, we are finding iron-sulfur clusters in the protein," Chazin says. It takes as many as seven genes for an organism to add an iron-sulfur cluster to a protein. And yet these genes—and the clusters—have been conserved by organisms from the dawn of life. "Nature doesn't put such a complicated functional group in there for no reason," Chazin says.

Not all repair proteins have iron-sulfur clusters. But when that group is missing or genes for DNA repair proteins are damaged, things can go terribly awry. Mutations in MutY's ability to transfer charges, for example, are associated with several hereditary forms of colon cancer. So Barton is now devoting about one-third of her lab to researching novel electronic cancer diagnostics capable of detecting how proteins commonly found in cancer affect DNA's conductivity, as well as medicines intended to seek out damaged DNA common in tumors.

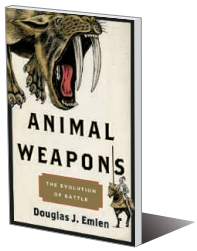
Even if medical uses of Barton's research don't pan out, her picture of DNA as a telephone wire would overhaul the biochemical narrative of life's most studied and most important molecule. "It would change our understanding of [DNA] transcription, replication and repair," Tainer says. That, he says, would be a discovery would be worthy of a Nobel Prize. "I'd nominate her for it." Barton dismisses the talk of prizes and says only that her exploration of DNA's wirelike behavior has been a wild ride. "It's remarkable. We started in one place and had no idea where it would take us," she says. ■

Published by AAAS

INSIGHTS

Environmental costs of animal drugs *p. 1296*

The evolution of animal weaponry *p. 1305*



PERSPECTIVES



Downloaded from www.sciencemag.org on December 14, 2014

PLANETARY SCIENCE

Scratching the surface of martian habitability

Are the right chemical ingredients enough to make a habitable planet?

By **Pamela G. Conrad**

Earth and Mars, though formed at the same time from the same materials, look very different today. Early in their histories they evolved through some of the same processes, but at some point their evolutionary paths diverged, sending them in perhaps irrevocably different directions. Knowledge of the factors that contributed to such different outcomes will help to determine how planets become habitable and how common habitable planets may be. The Mars surface environment is harsh today, but in situ measurements of ancient sedimentary rock by Mars Science Laboratory reveal chemical and mineralogical evidence of past conditions that might have been more favorable for life to exist. But chemistry is only part of what is required to make an environment habitable. Physical conditions constrain the chemical reactions that underlie life processes; the chemical and physical characteristics that make planets habitable are thus entangled.

Habitability can be defined by the requirements for Earth's simplest microbes, although it is conceivable that life that has evolved elsewhere may have different requirements. At a minimum, a habitable planet must have liquid water somewhere as an essential universal chemical for life. However, the thermal properties of water may be as important as its chemical properties. Its heat capacity and density as a solid play key roles in the evolution of habitable environments.

Today, Mars is cold and dry, with large diurnal fluctuations even at the equator. Ionizing radiation from galactic cosmic rays and solar ultraviolet (UV) light is sufficiently energetic to damage organic compounds. Life as we know it would be substantially challenged on the martian surface. Yet, prior to the late heavy bombardment (~4.1 to 3.8 billion years ago), the young Mars had an internal dynamo and magnetic field to shield itself from ionizing radiation and a thicker atmosphere to help moderate temperature swings and absorb harmful solar UV light. At that time, the martian surface may have been capable of sustaining microbial life within the range of known Earth extremophiles. It is a matter of timing: The emergence and sustenance of both life and its habitat require a convergence of the right chemicals and physical conditions in the right place at the same time.

In the mid-1970s, the two Viking landers with their Biological Experiment payloads (*1*) aimed to discover whether the chemical

NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA. E-mail: pamela.g.conrad@nasa.gov

PHOTO: NASA/JPL-CALTECH/MALIN SPACE SCIENCE SYSTEMS

Published by AAAS

constituents required for life were present and whether there was evidence that might indicate metabolism. In addition to seeking chemical evidence of life, the mission also measured aspects of the physical environment with its meteorology packages (*2*), thus setting a precedent for interrogating both the chemistry and the physics of the martian surface environment.

No consensus has been reached on whether the Viking mission detected signs of life. What Viking made clear, however, is that measurement of a variety of chemical and physical factors, on as large a number of samples as possible, is key to evaluating a complex problem such as the habitability of an environment.

Since then, rovers—including Pathfinder, the Mars Exploration Rovers, and Mars Science Laboratory (MSL)—have been able to do what Viking could not: drive to rocks that look like interesting targets. Yet, without the guidance of orbital imagery and spectroscopy to point the landers to those landing sites with the most scientific promise, exploration would proceed by trial and error. It was not a random walk that led Curiosity to the right place at Gale Crater to find promising chemical evidence of a formerly habitable environment (*3*–*6*). It took support from the two orbiters, Mars Reconnaissance Orbiter and Odyssey, to meet MSL's core science objective at Yellowknife Bay: to find evidence of an environment that is or was once habitable (*3*). The European Space Agency's ExoMars and NASA's Mars 2020 missions are now both selecting landing sites, guided by orbital imaging and spectroscopy and building on data returned from the Viking, Mars Exploration Rovers, Phoenix, and MSL missions.

The MAVEN (Mars Atmosphere and Volatile Evolution) mission arrived at Mars in September 2014 (*7*). It studies both the isotopic chemistry and some of the physical processes at the top of the martian atmosphere. Understanding these processes will be helpful when attempting to look back in time to extrapolate information about the original atmosphere and its implications for Martian habitability.

Since landing in 2013, MSL has regularly measured properties including radiation, ground and air temperature, atmospheric pressure, and wind speed and direction (*8*–*10*), providing a full year of data at the martian equator. Data provide information about the present dynamics of the surface environment, but some of the parameters are not easily preserved in the rock record, presenting a challenge to understanding the environmental physics of the past. Thus, even though the physical parameters are important for habitability, they are often cryptic in ancient samples. In situ radiometric and exposure age experiments on Mars now make it possible to date events such as changes in depositional environment in the martian rock record (*11*), and this will enable us to extrapolate rates of change over time. This is how we connect the contemporary measurements to a prediction of their past values. Dating milestones that mark the transition from a wetter and warmer Mars to the cold, dry and chemically oxidized Mars of today is within reach now that Mars can be explored with rovers that can go from exposed rocks of one system to those of another, drilling them and measuring their ages and time of exposure to the ionizing radiation at the surface.

Future missions to Mars will include even more complicated measurements than MSL's habitability assessment campaign. ESA's ExoMars mission intends to drill 2 m beneath the surface of Mars and seek chemical evidence for life. NASA's Mars 2020 mission will collect and cache samples with the goal of returning them to Earth for detailed interrogation in terrestrial laboratories. But as in situ missions to Mars become bolder and more technically capable, we must remain mindful about clues that reveal the physical attributes of ancient environments. To really understand whether Mars was once a habitable planet, it must be shown that it not only had the right chemistry, but the right physical conditions as well. ■



Was Mars ever habitable? Mastcam mosaic of an area called Rocknest in Gale Crater, Mars reveals a desolate environment. Image taken by Mars Science Laboratory's Curiosity rover in late 2012.

REFERENCES

1. H. P. Klein, *Icarus* **34**, 666 (1978).
2. S. L. Hess *et al.*, *Icarus* **16**, 196 (1972).
3. J. P. Grotzinger *et al.*, *Science* **343**, 1242777 (2014).
4. D. W. Ming *et al.*, *Science* **343**, 1245267 (2014).
5. S. M. McLennan *et al.*, *Science* **343**, 1244734 (2014).
6. D. T. Vaniman *et al.*, *Science* **343**, 1243480 (2014).
7. S. W. Bougher *et al.*, *Space Sci. Rev.* 10.1007/s11214-014-0053-7 (2014).
8. D. M. Hassler *et al.*, *Science* **343**, 1244797 (2014).
9. J. Gómez-Elvira *et al.*, *J. Geophys. Res. Planets* **119**, 1680 (2014).
10. R. M. Haberle *et al.*, *J. Geophys. Res. Planets* **119**, 440 (2014).
11. K. A. Farley *et al.*, *Science* **343**, 1247166 (2014).

10.1126/science.1259943

Published by AAAS

HIV

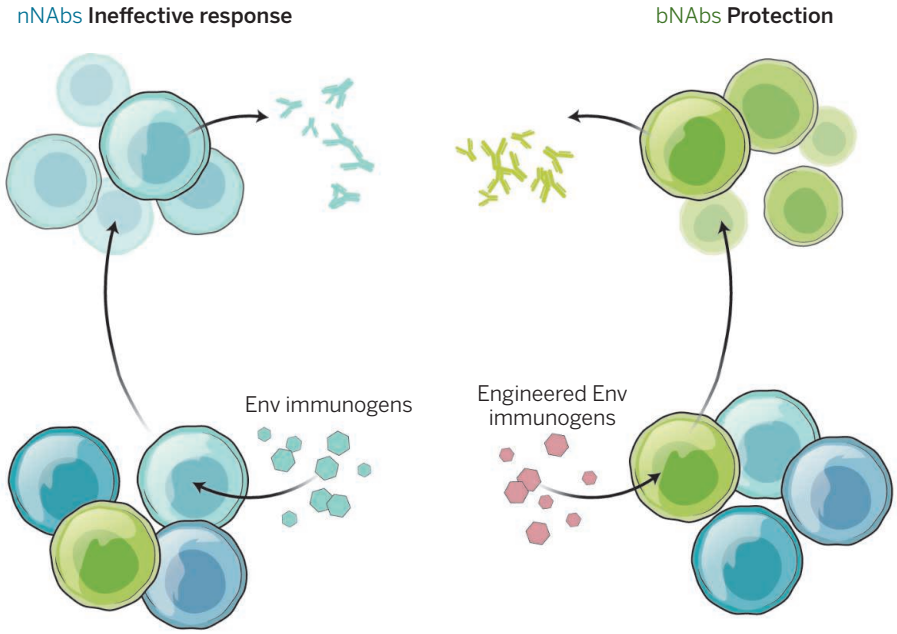
Expanding the breadth of an HIV-1 vaccine

Immunogens are designed to elicit the production of highly coveted broadly neutralizing antibodies that protect against HIV-1

By **Gilad Ofek**[1] *and* **Ron Diskin**[2]

One of the main problems encountered thus far in the quest for an effective HIV-1 vaccine has been that injected immunogens—virus-derived proteins against which an immune response is sought—generally only elicit antibodies with a narrow breadth of HIV-1 neutralization (*1*). During chronic infection, however, ~10 to 30% of individuals naturally develop highly efficient antibodies that can neutralize multiple HIV-1 viral strains. These broadly neutralizing antibodies (bNAbs) have provided a template for the design of immunogens aimed at eliciting similar antibodies upon vaccination. On page 1380 of this issue, McGuire *et al.* (*2*) show that immunogens derived from the viral envelope glycoprotein of HIV-1 preferentially activate B cells that produce only narrow neutralizing antibodies (nNAbs). This supersedes the activation of B cells that produce the desired bNAbs, when both are present in the same pool of cells and are exposed to the same immunogen, possibly explaining the inability of previous vaccine candidates to achieve protective immune responses. By modifying the immunogen, McGuire *et al.* show that this preferential activation can be reversed in favor of the desired B cells, offering exciting new promise for vaccine design.

A major recent breakthrough in the field of HIV-1 vaccinology has been the cloning of multiple highly potent bNAbs (*3*) and the discovery of their relative abundance in HIV-1–infected individuals. These bNAbs protected animals in virus challenge experiments (*4*, *5*) and helped treat ongoing infections in mice (*6*). Thus, eliciting similar bNAbs may be a key requirement for a successful HIV-1 vaccine. Of particular interest is a group of bNAbs that target a site on the surface of the trimeric spike complex of the viral envelope glycoprotein (Env) that binds to CD4, the HIV-1 receptor on host cells. Named VRC01-like monoclonal antibodies, after the first isolated member of this family (*7*), they include some of the most broad and potent bNAbs currently known (*8*). Notably, VRC01-like bNAbs have been isolated from different individuals and, due to similar structural constraints for recognition of the CD4-binding site, all have independently evolved to use the same germline gene that encodes the antibody heavy-chain subunit (*9*).



Better HIV-1 vaccine immunogens. Recombinant Env immunogens can be tested for the relative activation of B cells expressing either nNAb germline BCRs (shades of blue) or bNAb germline BCRs (shades of green) when both sets of B cells are present. Standard Env immunogens preferentially activate B cells expressing germline nNAb BCRs (left). Recombinant Envs can be engineered to preferentially activate B cells expressing bNAb germline BCRs as a first step in the eliciation pathway of mature bNAbs (right).

Elicitation of antibodies by the immune system, including VRC01-like antibodies, involves successful activation of B cells through engagement with foreign antigen. This recognition is mediated by B cell receptors (BCRs), which are antibody-like transmembrane proteins that appear as recombined germline versions on the surface of naïve B cells. Once activated, B cells proliferate and modify their BCR genes through somatic hypermutation. In this process, relative affinity for antigen preferentially selects B cells with the most fit BCR variants, leading to their secretion as mature antibodies. BCR activation also drives the uptake and processing of antigen for presentation to helper T cells, which may play an important role in the proliferation and maturation of B cells (*10*). An interesting property of VRC01-like bNAbs is the unusual extent to which they undergo somatic hypermutation. As a result, their germline versions do not bind Env-derived immunogens, posing the question of how best to design immunogens that can target this particular germline BCR and elicit a VRC01-like response.

Drawing on structural studies of VRC01-like antibodies in complex with gp120 (the receptor-binding subunit of HIV-1 Env), several approaches have been successfully undertaken in the design of Env molecules that target B cells carrying bNAb germ lines (*11*, *12*). The rationale for these designs has been to enable engagement of desired naïve BCRs upon immunization, thereby facilitating further hypermutation and elicitation of mature bNAbs. The recent determination of the crystal and electron microscopy structures of a soluble HIV-1 envelope trimer at near-atomic resolution (*13*, *14*), and the higher-resolution structure that was obtained (*15*), will likely aid this approach further by providing

[1]*Institute for Bioscience and Biotechnology Research, University of Maryland, Rockville, MD, USA.* [2]*Department of Structural Biology, Weizmann Institute of Science, Rehovot 76100, Israel. E-mail: ron.diskin@weizmann.ac.il*

ILLUSTRATION: V. ALTOUNIAN/*SCIENCE*

Downloaded from www.sciencemag.org on December 14, 2014

Published by AAAS

more accurate mimics of the native trimer that can be designed, as immunogens, to target the appropriate germ lines.

McGuire *et al.* compared the responses of B cells carrying germline versions of nNAbs and bNAbs to different Env immunogens. Using an assay that they developed, the authors found that a diverse array of nonengineered Envs can activate B cells that express germline nNAbs but not germline bNAbs, consistent with poor elicitation of bNAbs observed in previous vaccine studies. Envs designed to engage germline-bNAb BCRs (namely, 426c.NLGS.TM) nonetheless also activated B cells expressing germline-nNAb BCRs, and to a higher degree. Only when gp120 variable loops V1, V2, and V3 were deleted from the engineered Env did the ratio of activation reverse. Remarkably, germline-bNAb–expressing B cells were activated to a greater extent than those expressing germline-nNAb BCRs. These findings should guide the design of immunogens to identify new vaccine candidates that preferentially activate germline-bNAb–expressing B cells.

The in vitro approach developed by McGuire *et al.* may simplify certain aspects of B cell maturation that occur in lymphoid organs (germinal centers) after inoculation, including the suggested role of helper T cells. The extensive hypermutation of BCRs may not necessarily occur simply as a result of the disparate germline engagement of engineered Envs, and additional subsequent immunizations may be needed to drive the preferred maturation pathway. Further, when considering bNAbs that target other vulnerable sites on the Env trimer, which develop from a wider array of germline antibody genes, this approach may also have some limitations. However, from the perspective of vaccine design against the CD4-binding site, in which germline gene usage is critical, the study of McGuire *et al.* is an important advance in assessing vaccine immunogens for clinical development. ■

REFERENCES

1. L. E. McCoy, R. A. Weiss, *J. Exp. Med.* **210**, 209 (2013).
2. A. T. McGuire *et al.*, *Science* **346**, 1380 (2014).
3. F. Klein *et al.*, *Science* **341**, 1199 (2013).
4. M. Shingai *et al.*, *J. Exp. Med.* **211**, 2061 (2014).
5. A. B. Balazs *et al.*, *Nature* **481**, 81 (2012).
6. F. Klein *et al.*, *Nature* **492**, 118 (2012).
7. X. Wu *et al.*, *Science* **329**, 856 (2010).
8. J. F. Scheid *et al.*, *Science* **333**, 1633 (2011).
9. A. L. West, Jr *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 5633 (2012).
10. G. D. Victora, M. C. Nussenzweig, *Annu. Rev. Immunol.* **30**, 429 (2012).
11. J. Jardine *et al.*, *Science* **340**, 711 (2013).
12. A. T. McGuire *et al.*, *J. Exp. Med.* **210**, 655 (2013).
13. D. Lyumkis *et al.*, *Science* **342**, 1484 (2013).
14. J.-P. Julien *et al.*, *Science* **342**, 1477 (2013).
15. M. Pancera *et al.*, *Nature* **514**, 455 (2014).

10.1126/science.aaa2315

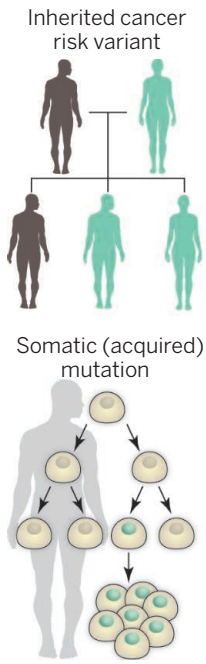
CANCER

Cancer by super-enhancer

Tiny changes in our genomes can enhance oncogene expression and contribute to tumorigenesis

By **Anna Vähärautio**[1] *and* **Jussi Taipale**[1,2]

Most recurring somatic mutations in cancer affect protein-coding regions, either through activating oncogenes or inactivating tumor suppressors (*1*). However, several classes of mutations have been identified that affect the much larger noncoding regions of the genome, leading to changes in gene expression. These include large-scale genomic rearrangements that bring a strongly active promoter next to an oncogene, or place oncogene promoters under the influence of strong transcriptional enhancers (see the figure, panel A). On page 1373 of this issue, Mansour *et al.* (*2*) identify a very small mutation in ~5% of T cell acute lymphoblastic leukemias (T-ALL; see the figure, panel B) that generates a large, powerful enhancer capable of driving tumorigenesis.



Genetics of cancer. Both inherited variants (**top**) and acquired mutations (**bottom**) can contribute to tumorigenesis.

The recent discovery that most inherited genomic variants that predispose us to disease reside in the noncoding portion of our genomes has generated considerable interest in mapping somatic (acquired) mutations in such gene regulatory regions of cancer cells. One such effort has led to the identification of causative, "driver" mutations affecting the proximal promoter of telomerase catalytic subunit (TERT) in melanoma (*3*, *4*). Despite the rapidly expanding list of causative inherited variants affecting distal enhancers (*5*), somatic driver mutations affecting enhancers had not until now been identified.

The discovery by Mansour *et al.* was based on the observation that a subset of T-ALL expressed high levels of the *TAL1* oncogene, a DNA-binding transcription factor, but did not contain the previously identified large rearrangements near *TAL1* that could explain the high expression level. The lymphoma cells expressed only one of the two copies (alleles) of the *TAL1* gene, suggesting the existence of a novel "cis-acting" mutation near the overexpressed copy. To identify the site of the mutation, Mansour *et al.* studied whether the surrounding regions had gene regulatory potential. They found that two T-ALL cell lines, Jurkat and MOLT-3, harbored a large 20-kb region with high levels of a histone modification associated with transcriptional enhancer and promoter activity (acetylated histone H3 lysine 27). Sequencing of this region from the cell lines and from primary T-ALL revealed small insertions of 2 to 18 base pairs (bp) that were located ~7.5 kb upstream of the *TAL1* transcription start site. All the insertions introduced one or two de novo recognition sites for transcription factors of the MYB (myeloblastosis) family. Further experiments indicated that these sites were indeed bound by MYB. MYB binding also recruited to the affected region several other transcription factors that regulate hematopoiesis (GATA3, TCF12, RUNX1, and TAL1), as well as the transcriptional coactivator CBP.

These results suggested that the mutations had generated a novel enhancer that was responsible for the allele-specific overexpression of TAL1. To test this hypothesis, Mansour *et al.* set out to study the requirement of the MYB sites using CRISPR–Cas9 genome editing methods. Specific deletion of the enhancer region resulted in loss of both the histone acetylation signal and TAL1 expression, indicating that the observed enhancer-forming mutation functions as a bona fide driver mutation in T-ALL. Furthermore, the hematopoietic transcription factors that bind to the TAL1 enhancer are known to cooperate with the TAL1 gene product, and regulate each other, forming a self-reinforcing autoregulatory loop. The gain of the MYB site(s) upstream of the *TAL1* gene may thus act by completing the loop, and generating a pathological "locked" tran-

[1]*Genome-Scale Biology Program, University of Helsinki, Finland.* [2]*Science for Life Laboratory, Department of Biosciences and Nutrition, Karolinska Institutet, Stockholm, Sweden. E-mail: jussi.taipale@ki.se*

ILLUSTRATION: P. HUEY/*SCIENCE*

Downloaded from www.sciencemag.org on December 14, 2014

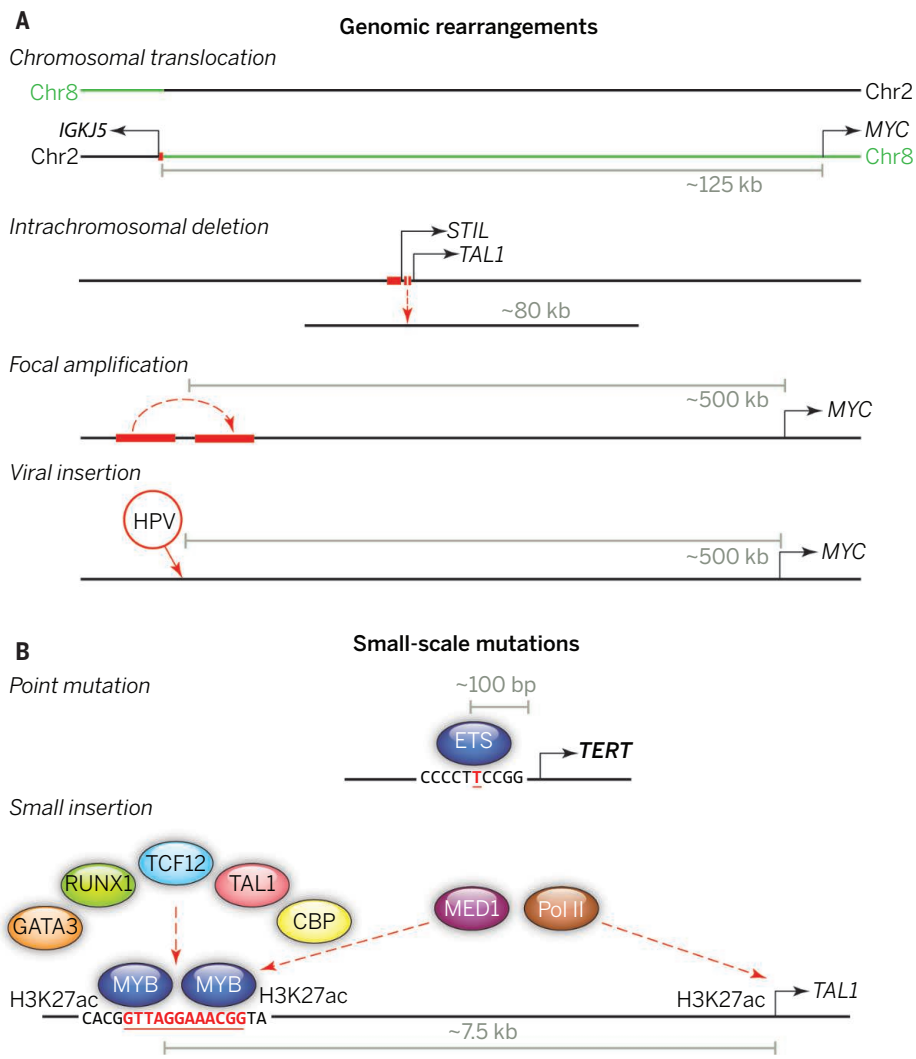
Published by AAAS

scriptional state that drives tumorigenesis. Given that similar regulatory interactions have also been found in T-ALL cases that carry different upstream lesions, it is tempting to speculate that the specific collection of driver mutations present in each individual tumor invariably results in formation of such locked transcriptional states that drive proliferation.

In contrast to the previously observed enhancer-linked driver mutations that are the result of massive genome rearrangements (see the figure), Mansour *et al.* show that even a small insertion of a few base pairs that introduces only one or two binding sites for a transcription factor can generate a large active enhancer capable of driving tumorigenesis. This exemplifies the magnitude of the challenge that lies ahead: In addition to the function of the reference genome, it is necessary to understand the regulatory activity of the unique genomes of each individual tumor. The discovery by Mansour *et al.* will also undoubtedly inspire studies aimed at identifying similar lesions in other tumor types. However, unbiased identification of small driver mutations in noncoding sequence is challenging because of our limited understanding of the gene regulatory code, and the difficulty of unequivocally identifying small insertions and deletions using short-read sequencing. Given these challenges, the targeted approach used by Mansour *et al.*—focusing on allele-specific expression and mutation-induced regulatory activity, followed by rigorous loss-of-function analysis—is especially promising.

The study by Mansour *et al.* is also notable in that it provides an example of the functional importance of genomic features that carry very strong signals for enhancer-associated chromatin marks. Such regions have been variably identified as locus control regions (*6*), DNase I super-hypersensitive sites (*7*), or more recently, super-enhancers (*8*). At present, there is no consensus on a strict definition of such regions. In addition, it is unclear whether such elements exist as a distinct biochemical entity, or whether differences between them and "common" enhancers are merely quantitative rather than qualitative. Genomics has traditionally focused on the more qualitative aspects of genomes, identifying and mapping features such as genes, transcripts, binding sites, and histone modifications. Less emphasis has been placed on the relative concentration, activity, or specific importance of the identified features. But elementary biochemistry teaches us that high-order reactions involving a large number of cooperative interactions are particularly sensitive to changes in reactant concentrations. Consistently, several studies have shown that genes that have strong enhancers are more sensitive to perturbations that decrease the activity of key transcriptional regulators, such as cohesin, mediator, chromatin modifiers (for example, BRD4), and components of the transcriptional machinery (for example, CDK7) (*8*–*10*).

Many cancer driver genes are brought under the influence of very strong enhancers during tumorigenesis (see the figure). Thus, after locating enhancer-associated genomic landmarks, researchers should pay close attention to their characteristics. Especially as super-enhancers, associated with the most extensive enhancer signatures, are particularly sensitive to perturbation, and may thus become a promising new target of antineoplastic agents. ■



Noncoding mutations in cancer. (**A**) Schematic representation of large-scale genomic rearrangements. Interchromosomal translocations (top) can result in placement of oncogene promoter (e.g., *MYC*) under the influence of a strong regulatory region (e.g., those from Igκ locus, red). Intrachromosomal deletions can similarly lead to increased expression of oncogenes, exemplified by deletion of an 80-kb sequence in chromosome 1, leading to expression of the coding sequence of the *TAL1* oncogene under the broadly active STIL1 promoter (*11*). Regions with transcriptional enhancer activity can also be affected by focal amplifications (*12*) and viral insertions (*13*). (**B**) Small-scale somatic driver mutations affecting noncoding sequence. (Top) Base substitutions at the proximal promoter of the *TERT* oncogene create recognition sequences for ETS transcription factors (*3*, *4*). (Bottom) Mansour *et al.* report noncoding insertion mutations that create binding sites for the MYB transcription factor, and lead to formation of a de novo enhancer that drives expression of the *TAL1* oncogene.

REFERENCES

1. B. Vogelstein *et al.*, *Science* **339**, 1546 (2013).
2. M. R. Mansour *et al.*, *Science* **346**, 1373 (2014).
3. S. Horn *et al.*, *Science* **339**, 959 (2013).
4. F. W. Huang *et al.*, *Science* **339**, 957 (2013).
5. I. Sur *et al.*, *Cancer Res.* **73**, 4180 (2013).
6. F. Grosveld *et al.*, *Cell* **51**, 975 (1987).
7. T. M. Ryan *et al.*, *Genes Dev.* **3**, 314 (1989).
8. J. Lovén *et al.*, *Cell* **153**, 320 (2013).
9. J. M. Rhodes *et al.*, *Dev. Biol.* **344**, 637 (2010).
10. N. Kwiatkowski *et al.*, *Nature* **511**, 616 (2014).
11. F. Mitelman *et al.*, *Nat. Rev. Cancer* **7**, 233 (2007).
12. J. Edelmann *et al.*, *Blood* **120**, 4783 (2012).
13. A. Adey *et al.*, *Nature* **500**, 207 (2013).

10.1126/science.aaa3247

ILLUSTRATION: P. HUEY/*SCIENCE*

Published by AAAS

ultraviolet (XUV) absorption spectroscopy to distinguish between electronic and lattice contributions. A short-wavelength probe photon can only be absorbed if the energy exceeds the gap between the well-defined core level and the initially free conduction band. Populating the conduction band with the optical pulse will lower the x-ray absorption probability because the band is now partially filled.

Additional information about the electronic response can be obtained from the observation of a shift of the gap between the valence and conduction band, manifested in a shift of the x-ray absorption edge. However, x-ray absorption spectroscopy can provide even more information by absorbing a photon energetic enough to ionize the atom. The outgoing photoelectron will interact with the neighboring atoms and modifies the absorption probability depending on the energy. The absorption spectrum is described within the framework of XANES (x-ray absorption near-edge structure) (*8*) and provides specific information about the local environment of the atom perturbed by the pump pulse. Time-resolved x-ray absorption spectroscopy with femto- and attosecond resolution offers a much more detailed picture about the electronic and lattice dynamics of an excited semiconductor than will be possible with pure optical probing (*9*, *10*).

With this powerful spectroscopic tool, Schultze *et al.* studied the intensity-dependent excitation dynamics of electrons in silicon in the tunneling regime. For high intensities, the density of electrons in the conduction band increased in a steplike manner because tunneling excitation was restricted to the small temporal window near the pulse maximum, which is in excellent agreement with their theoretical predictions (see the figure). This work begins a new era in solid-state spectroscopy, showing not only unprecedented temporal resolution but also a reliable and intuitive method to disentangle the different mechanisms connected with light absorption in a solid. ■

REFERENCES

1. A. Einstein, *Ann. Phys.* **17**, 132 (1905).
2. L. V. Keldysh, *Sov. Phys. JETP* **20**, 1307 (1965).
3. M. Hentschel *et al.*, *Nature* **414**, 509 (2001).
4. M. Uiberacker *et al.*, *Nature* **446**, 627 (2007).
5. M. Schultze *et al.*, *Science* **346**, 1348 (2014).
6. A. Schiffrin *et al.*, *Nature* **493**, 70 (2013).
7. J. Shah, *Ultrafast Spectroscopy of Semiconductors and Semiconductor Nanostructures*, *Springer Series in Solid-State Sciences*, Vol. 115 (Springer-Verlag, Berlin, ed. 2, 1999).
8. D. C. Koningsberger, R. Prins, *X-ray Absorption: Principles, Applications, Techniques of EXAFS, SEXAFS, and XANES* (Wiley, New York, 1988).
9. S. L. Johnson *et al.*, *Phys. Rev. Lett.* **91**, 157403 (2003).
10. E. Seres, C. Spielmann, *Appl. Phys. Lett.* **91**, 121919 (2007).

10.1126/science.aaa1211

EPIDEMIOLOGY

Mathematical models for emerging disease

How can models provide better insights into pathogen emergence and epidemic prediction?

By **Andy Dobson**[1,2]

It has been nearly 25 years since the publication of *Infectious Disease of Humans* (*1*), the "vade mecum" of mathematical modeling of infectious disease; the proliferation of epidemiological careers that it initiated is now in its fourth generation. Epidemiological models have proved very powerful in shaping health policy discussions. The complex interactions that lead to pathogen (and pest) outbreaks make it necessary to use models to provide quantitative insights into the counterintuitive outcomes that are the rule of most nonlinear systems. Thus, epidemic models are most interesting when they suggest unexpected outcomes; they are most powerful when they describe the conditions that delineate the worst-case unexpected scenario, and provide a framework in which to compare alternative control strategies. But what are the limits of mathematical models and what kinds provide insight into emerging disease?

Most mathematical models for infectious disease initially develop a framework to estimate R_0, the basic reproductive number of the disease, or the average number of secondary cases produced by the first infected individual (*1*). R_0 is the central concept of mathematical disease modeling. Its utility can be illustrated with a simple model for the current Ebola epidemic (see the first figure). The model suggests that the numbers of people dying and how quickly the pathogen is eradicated have a nonlinear relationship with R_0 and the speed and efficiency at which those exposed to Ebola are identified and quarantined (see the second figure). Thus, where resources are limited for quarantine and contact tracing (finding everyone who comes in direct contact with a sick Ebola patient), large epidemics quickly result. As quarantine delays get longer, R_0 increases to its peak value. Once R_0 is bigger than 1, the numbers of people dying in the epidemics rise quickly. Plainly, the key intervention is to quarantine people as quickly as possible, preferably faster than the average time to develop symptoms. Delays in quarantine quickly increase the final size of the epidemic. From a control perspective, the key variables to be quantified are the magnitude of R_0 and the length of time in which infected individuals transmit the pathogen before symptoms appear (*2*); in the case of Ebola, several days may elapse before symptoms appear, so it is essential to quickly isolate contacts to control outbreaks (*3*). Notably, this was not the case with HIV (and tuberculosis), where people were infectious for years and showed no symptoms; ultimately this allowed HIV to spread globally (*4*).

By using an initial estimate for R_0 and a set of equations that describe the etiology of a pathogen, a quantitative framework can be developed for population-level interventions such as vaccination. Work on measles forms the central core of this area of infectious disease modeling; it is based on the detailed, spatially explicit, time series of spatially located cases of measles available for towns and cities in the United

Quarantine, R_0, and the epidemic dynamics of Ebola

$$R_0 = \frac{\beta N \varepsilon}{(\varepsilon + q)(\delta + \alpha + q)}$$

The efficacy of tracing quarantine can be examined with a simple model in which the host population (N) is divided into susceptible, exposed, infectious, quarantined, and recently deceased individuals whose dynamics can be described by an expression for R_0 [parameters of the model can be estimated from the current epidemic (*3*)]. β, transmission rate; ε, rate of host infectiousness; q, rate of movement into quarantine; δ, rate of sick recovery; α, mortality rate of infected hosts. See the second figure and supplementary materials for details of the equation.

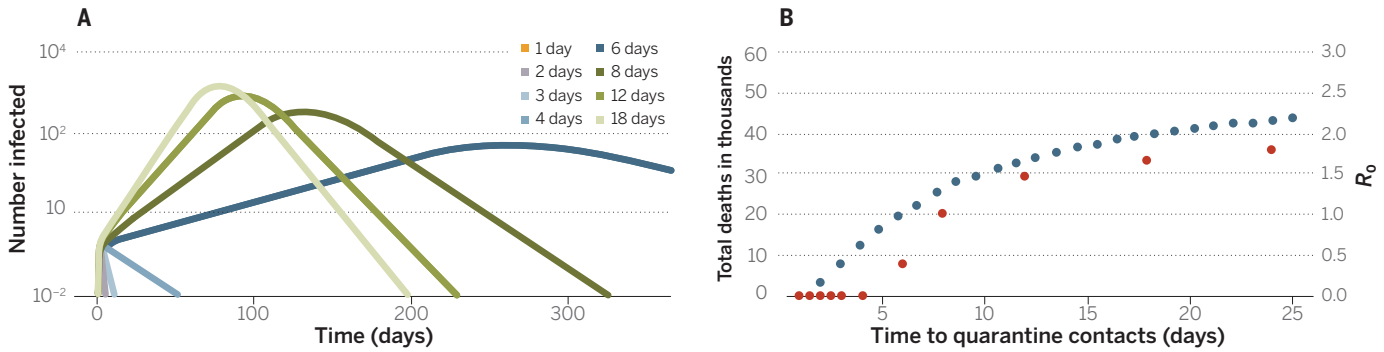
[1]*Department of Ecology and Evolutionary Biology (EEB), Eno Hall, Princeton University, Princeton, NJ 08544, USA.* [2]*Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA. E-mail: dobson@princeton.edu*

Downloaded from www.sciencemag.org on December 14, 2014

Corrected 11 December, 2014; see full text.

Published by AAAS

Quarantine delays and size of Ebola outbreak



Ebola dynamics. (**A**) Dynamics of the epidemic in a West African town of 500,000 people for different levels of quarantine efficiency: 1, 1.5, 2, 3, 4, 6, 8, 16, 24 days. (**B**) The magnitude of R_0 (blue circles) and the total number of people dying in the first year of the Ebola outbreak in a hypothetical town of 50,000 people (red circles). Graphs were generated applying the model shown in the first figure and data from (*3*).

Kingdom (*5*, *6*). The measles approach can be extended to pathogens whose dynamics are coupled to a strong climate signal—for example, cholera (*7*) and malaria (*8*). Here, the models can be coupled to long-term climate data and climate predictions, to provide outbreak forecasts many months into the future. Work in this area has given rise to powerful new inferential fitting methods for models that considerably increase the accuracy of outbreak predictions (*9*), while also allowing models whose structure reflects different underlying assumptions to be compared. These approaches move well beyond time series and statistical regression analyses as they include mechanistic details as mathematical functions that define rates of loss of immunity and the response of vector abundance to climate.

By contrast, in the absence of a time series of data for prior outbreaks, predicting outbreaks for novel or emerging pathogens has been marked by lack of accuracy (*10*). The problem is directly analogous to predicting hostile cyber-attacks on computer systems; in both situations, the quality of useful information in the early stages of an outbreak is limited and swamped by a diversity of other signals leading to too many false-positive predictions. By extension, this problem is hugely compounded with maps that seek to illustrate epidemic hot-spots for emergent disease (*11*). These are at best misleading (*10*). Many of those creating risk maps argue that new genetic methods will allow the identification of viruses and bacteria that will become emerging diseases. Recent work on determinants of disease severity for closely related parasite genotypes of rodents suggest that there are substantial hidden challenges to this approach. For example, *Plasmodium chabaudi*, a species of malaria that infects rodents, has a variety of laboratory strains that are genetically very similar (*12*); they were produced by diluting field isolates and then seeing which one would grow in lab strains of mice. The pathogen strains are almost genetically identical but have very different properties when injected into lab mice (*13*)—some die out, other proliferate robustly; still others proliferate modestly yet kill the mice at relatively low levels of infection. Although emergent viral pathogens have much smaller genomes than *Plasmodium*, if we cannot yet predict what a particular genome will do in a well-studied lab model, then it is unclear that we can predict whether a recently discovered virus or bacteria will be pathogenic, benign, or initially innocuous and ultimately fatal.

Two developing approaches may eventually help in this effort. The techniques used by ecologists to predict sudden changes of state in lakes and other ecosystems rely on time series that can quantify when the variance of the signal changes rapidly (*14*). This may be possible for pathogens, but will still be inherently dependent upon long-term data to seed the predictive algorithm. Alternatively, if prior information about spatial distribution and contact rates between hosts is available, it may be possible to make spatial and temporal predictions for outbreaks, particularly if basic data on incubation and infectious period of infected hosts and the efficacy of pathogen transmission can be quickly obtained.

Ultimately, emergence and prediction are very different phenomena that have an asymmetrical and complex interaction with each other. Complex ecological and epidemiological systems have emergent properties, but the ability to predict these properties, or even their emergence, will be very limited until retrospective data are available to which some form of structural model can be fitted.

Mathematical models are the only way to formally test the interaction between what is known and what is only partially understood about any evolving or emerging pathogen and its host population. Although recent reviews (*15*) provide some key insights (e.g., most new human infections will be zoonotic viruses), the details needed for public health intervention (what, where, and how many cases) are inherently unpredictable. For that reason, better surveillance for pathogens and more rapid response capacity for disease outbreaks are needed. Hopefully, the current Ebola crisis will catalyze substantial changes in the ways we monitor and respond to future emergent disease outbreaks. ■

REFERENCES AND NOTES

1. R. M. Anderson, R. May, *Infectious Diseases of Humans* (Oxford Univ. Press, Oxford, 1992).
2. C. Fraser et al., *Proc. Natl. Acad. Sci. U.S.A.* **101**, 6146 (2004).
3. W. E. R. Team, *N. Engl. J. Med.* **371**, 1481 (2014).
4. R. M. Anderson, D. R. Cox, H. C. Hillier, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **325**, 39 (1989).
5. B. T. Grenfell et al., *Ecol. Monogr.* **72**, 185 (2002).
6. B. T. Grenfell, B. M. Bolker, *Ecol. Lett.* **1**, 63 (1998).
7. K. Koelle et al., *Nature* **436**, 696 (2005).
8. K. Laneri *et al.*, *PLOS Comput. Biol.* **6**, e1000898 (2010).
9. E. L. Ionides, C. Bretó, A. A. King, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 18438 (2006).
10. A. P. Dobson, in *The Influence of Global Environmental Change on Infectious Disease Dynamics Forum on Microbial Threats*, E. Choffnes, A. Mack, Eds. (National Academies Press, Washington, DC, 2014), chap. A8, pp. 193–212.
11. K. E. Jones *et al.*, *Nature* **451**, 990 (2008).
12. D. Walliker, R. Carter, S. Morgan, *Nature* **232**, 561 (1971).
13. G. H. Long *et al.*, *BMC Evol. Biol.* **8**, 128 (2008).
14. M. Scheffer *et al.*, *Science* **338**, 344 (2012).
15. J. O. Lloyd-Smith *et al.*, *Science* **326**, 1362 (2009).

ACKNOWLEDGMENTS

I thank P. Farmer for suggesting examining quarantine and Ebola as an indication of how poverty has generated the Ebola epidemic in West Africa. A discussion that S. Levin organized at EEB, Princeton, on "Emergence" provided several key insights from A. Graham and S. Pacala that I have gratefully used here. My work on infectious diseases is sponsored by the NSF-NIH Ecology Infectious Disease program, Princeton Grand Challenges, and by the James McDonnell Fund.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/346/6215/1294/suppl/DC1
Supplementary Text

10.1126/science.aaa3441

ILLUSTRATION: G. GARVIN/*SCIENCE*

Corrected 11 December, 2014; see full text.

Published by AAAS



Griffon Vultures at a feeding station in Lleida, Spain.

SCIENCE AND REGULATION

One Health approach to use of veterinary pharmaceuticals

Weak environmental assessments undermine regulations

By **A. Margalida*, G. Bogliani, C. G. R. Bowden, J. A. Donázar, F. Genero, M. Gilbert, W. B. Karesh, R. Kock, J. Lubroth, X. Manteca, V. Naidoo, A. Neimanis, J. A. Sánchez-Zapata, M. A. Taggart, J. Vaarten, L. Yon, T. Kuiken†, R. E. Green**

POLICY

An estimated 6051 tons of active substances went into the production of veterinary pharmaceuticals (VPs) for the treatment of food animals in the European Union (EU) in 2004, including 5393 tons of antibiotics and 194 tons of antiparasitics (*1*). With global meat production projected to increase (*2*) and the growing market for companion animal pharmaceuticals (*3*), the use of VPs will continue to increase. Although VPs may benefit the health and welfare of domestic animals and the efficiency of food animal production, they can contaminate the environment through manufacturing, treatment of animals, and disposal of carcasses, offal, urine, feces, and unused products (*4*) (see the chart). This contamination is a threat to nontarget species, including humans. With Spain having recently authorized marketing of a VP that was banned in South Asia in the past decade in light of environmental impacts, we recommend strengthening of current procedures and addition of a more proactive, holistic, One Health approach applicable to all VPs.

VULNERABLE VULTURES. In the 1980s, the three *Gyps* vulture species endemic to South Asia were the most abundant large raptors in the world, but their populations were reduced to near extinction in the 1990s (*5*). The nonsteroidal anti-inflammatory drug (NSAID) diclofenac was identified (in 2004) as the primary cause of rapid declines in Pakistan, India, and Nepal. Low-cost diclofenac-based products were being widely administered to livestock. Sufficient residues remained in carcasses of treated animals to cause acute renal failure and death of vultures feeding on them. Lethal contamination of just 0.3 to 0.7% of ungulate carcasses could account for observed decline of one vulture species at 50% per year (*6*).

The government of India enacted a ban on production, importation, and sale of veterinary diclofenac products in 2006. Similar measures were taken in Pakistan, Nepal, and Bangladesh. This was facilitated by the identification of meloxicam as a suitable alternative drug that was safe for *Gyps* vultures. Over the past 8 years, vulture population declines in South Asia have slowed, and may have reversed in some areas (supplementary materials) (*7*, *8*).

Despite this history, the government of Spain authorized marketing of diclofenac as a VP for use in cattle, pigs, and horses in 2013. This authorization was in compliance with current EU guidelines. Spain is important for global conservation of avian scavengers, as it holds >95% of the European population of vultures, the entire population of the globally threatened Spanish Imperial Eagle (*Aquila adalberti*), and important numbers of Red Kites (*Milvus milvus*) (table S1 and fig. S1).

Spanish law allows carcasses of farm animals to be left in the field in some protected areas or to be taken to "muladares" (vulture feeding stations) to provide food for wildlife. By law, diclofenac should only be administered under veterinary supervision and should not be given to animals that are likely to enter the natural food chain. But Spain's livestock industry has around 25 million pigs and 5.7 million cattle, and diclofenac is licensed for use against many clinical conditions that occur in these animals. There has been a dramatic increase in veterinary use of NSAIDs in recent decades (*9*). Vulture populations are very sensitive to even very low levels of contamination. Thus, despite existing regulations, given the scale of use and the reality of imperfect compliance with regulations, it seems reasonable that diclofenac could still find its way into the vulture food chain, with potentially harmful outcomes.

Vultures have traditionally provided important ecosystem services, helping control disease and pests, recycling nutrients, and providing cultural inspiration and recreational value. It has been estimated that Spanish vultures remove >8000 tons of livestock carcasses per year alone, preventing release of greenhouse gases and providing economic savings estimated at €1.5 million ($1.86 million) (*10*).

European countries have important populations of other endangered avian scavengers, and these depend heavily on livestock carcasses in some areas (*11*). Consequences of use of NSAIDs are likely to occur beyond the borders of individual countries, as many species show pronounced seasonal and erratic movements (*12*) (fig. S2). The toxicity of diclofenac to most accipitriforms is largely unknown, but an eagle species, *Aquila nipalensis*, may be susceptible (*13*). The risk to avian scavengers has not been evaluated adequately, thus diclofenac should be suspended for veterinary use in the EU.

In response to concerns raised by mem-

*See supplementary materials for author affiliations.
†t.kuiken@erasmusmc.nl

PHOTO: JORDI BAS

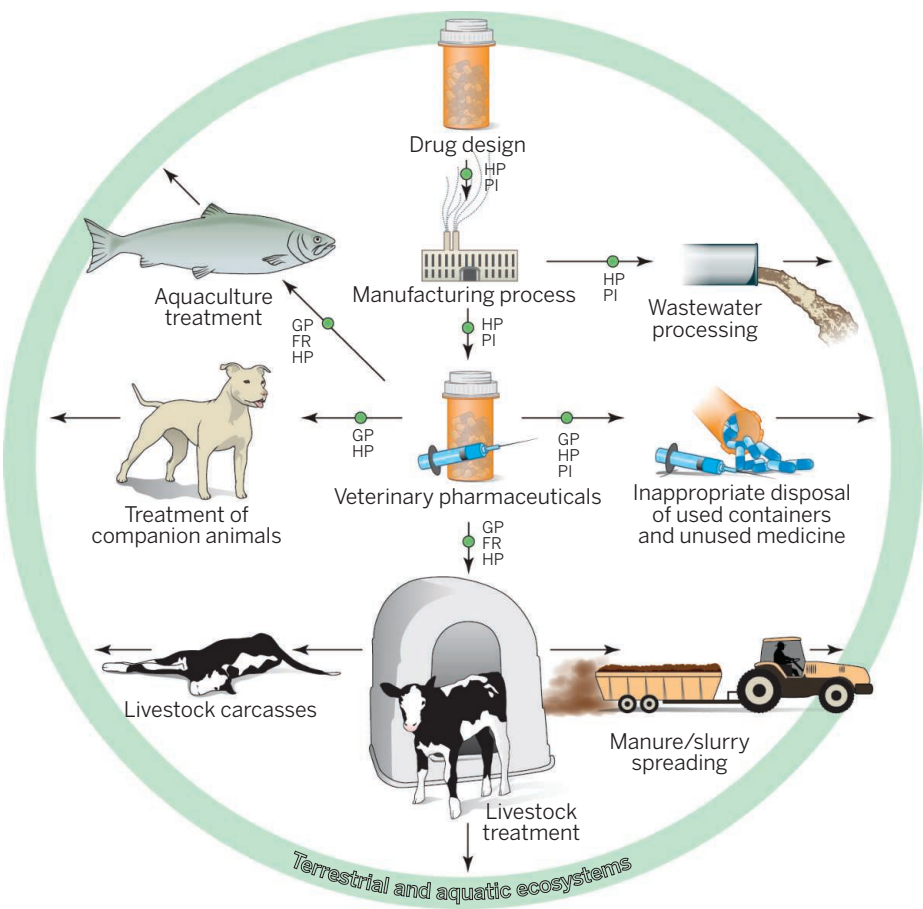
Downloaded from www.sciencemag.org on December 14, 2014

Published by AAAS

bers of the public, politicians, and conservation organizations, the European Commission will consider scientific advice on possible effects of veterinary medicines containing diclofenac on avian scavengers. This may lead to withdrawal of diclofenac for veterinary use in the EU and, it is hoped, convince other countries to follow.

A FLAWED APPROACH. Environmental risk assessment for new VPs is necessary for national licensing in EU countries. But there are approximately 2000 VPs in use in the EU, most of which have never been fully tested (*14*). VPs are exempt from assessment if they are used in a nonfood species, in a minor food species (i.e., all species except cattle, pigs, chickens, sheep for meat, and Atlantic salmon) if reared the same way as a major species for which an assessment already exists, or belong to certain product types (considered to be used for "a small number of animals in a herd or flock"): anesthetics; sedatives; injectable antibiotics (except those used for pigs, respiratory disease in cattle, or foot rot in sheep); injectable corticosteroids; hormones (except those that have a zootechnical use); and injectable NSAIDs (*15*). Assessment does not account for several key issues, many of which we remain largely ignorant of, including effects on species other than the few tested, low-dose effects, chronic effects, interactive effects after exposure to multiple pharmaceuticals, exposure during vulnerable stages (such as gestation and development), rate of degradation of pharmaceuticals, and toxicity of metabolites (*4*).

Diclofenac would be exempt from environmental impact assessment because it is an injectable NSAID, despite its known toxicity in nontarget species such as vultures. Other NSAIDs used in the EU may also pose a risk to avian scavengers. Ketoprofen is nephrotoxic in African *Gyps* vultures at doses likely to be encountered when feeding on carcasses of ungulates given a standard veterinary dose (*16*), and carprofen and flunixin may also be nephrotoxic to *Gyps* vultures (*5*, *17*). NSAIDs are one of several categories of VPs that may pose a risk to nontarget species through environmental contamination. Others include parasiticides and their nontarget impact on vertebrates and invertebrates in dung, soil, and watercourses (*18*); and more broadly, the human health implications of antimicrobial resistance in environmental bacteria associated with large-scale antibiotic use in food animal production (*19*).



Major pathways of release of veterinary pharmaceuticals into the environment. Green dots represent control points where environmental contamination can be prevented or minimized. GP, general public; FR, food retailers; HP, health professionals and scientists, including veterinarians, pharmacologists, farmers, animal scientists, ecologists, and environmental scientists; PI, pharmaceutical industry. Based in part on (*4*).

ILLUSTRATION: P. HUEY/*SCIENCE*

A ONE HEALTH MINDSET. We may never have the knowledge required to adequately assess environmental risk of VPs. Whereas we need to strengthen current systems of environmental impact assessment where possible, we also need to foster the precautionary principle and aim to prevent environmental contamination with VPs in the first place. We advocate "cradle-to-cradle" stewardship that promotes environmental responsibility, involves all sectors of society and considers environmental effects during production, use, and disposal (see the chart).

General public. Increase public education regarding environmental effects of pharmaceuticals and personal care products [e.g., a brochure (www.vulpro.com) for South African farmers and land owners, explaining risks of VP-contaminated carcasses]; promote take-back programs at pharmacies and veterinary clinics to facilitate appropriate disposal of unused medication [patients informed of environmental consequences of pharmaceuticals were more likely to return unused medicines for proper disposal (*20*)]; use consumer purchasing power to encourage environmentally sustainable food animal production (e.g., eating-better.org).

Food retailers and restaurants. Source and promote food products of animal origin that are generated under environmentally sustainable conditions. Organizations like sustainweb.org give practical guidelines for restaurants and caterers to adopt a sustainable approach to food.

Professionals and scientists. Stimulate collaboration among veterinarians, pharmacologists, farmers, animal scientists, ecologists, and environmental scientists, who are often unaware of each other's work. These disciplines are integral for design, dispensing, and application of VPs and for creating animal husbandry systems that promote biosecurity and principles of hygiene and that contribute to the health of food animals and their consumers, while maintaining the integrity of the environment and safety for nontarget species. The Swedish Environmental Classification and Information System for Pharmaceuticals brings together the pharmaceutical industry, Swedish Medical Products Agency, regional authorities, and physicians to provide tools for prescribing drugs in an environmentally conscious way (*21*).

Pharmaceutical industry. Practice "green pharmacy" (*22*) by considering the environment at all stages of the pipeline: drug de-

Published by AAAS

sign, delivery, packaging, advertising, and marketing. The pharmaceutical industry already has moved toward products and processes that are more environmentally sustainable, e.g., by use of enzymes for some transformation reactions and use of continuous processes for primary and secondary pharmaceutical production (*23*).

Such stewardship for VPs would mirror similar programs proposed for human pharmaceuticals and personal care products (*22*, *24*) and would help restrict effects of pharmaceuticals to where they belong: in the target species. This integrated effort to link the health of people, animals, and the environment is a good example of the One Health approach, an important step toward a more sustainable society. ■

REFERENCES AND NOTES

1. S. A. Kools, J. F. Moltmann, T. Knacker, *Regul. Toxicol. Pharmacol.* **50**, 59 (2008).
2. N. Alexandratos, J. Bruinsma, "World agriculture towards 2030/2050, the 2012 revision" (FAO Agricultural Development Economics Division, Rome, 2012).
3. National Office of Animal Health, Facts and figures about the UK animal medicines industry;www.noah.co.uk/focus/facts_figures.htm.
4. A. B. Boxall, *Handbook Exp. Pharmacol.* **199**, 291 (2010).
5. D. J. Pain *et al.*, *Bird Conserv. Int.* **18** (S1), S30 (2008).
6. R. E. Green *et al.*, *J. Appl. Ecol.* **41**, 793 (2004).
7. M. J. I. Chaudhry *et al.*, *Bird Conserv. Int.* **22**, 389 (2012).
8. V. Prakash *et al.*, *PLOS ONE* **7**, e49118 (2012).
9. G. M. Friton, C. Cajal, R. Ramirez-Romero, *Vet. Rec.* **156**, 809 (2005).
10. A. Margalida, M. A. Colomer, *Sci. Rep.* **2**, 753 (2012).
11. J. A. Sánchez-Zapata *et al.*, *Bird Study* **57**, 352 (2010).
12. J. M. Thiollay, in *Handbook of the Birds of the World*, J. del Hoyo, A. Elliott, J. Sargatal, Eds. (Lynx Edicions, Barcelona, 1994), vol. 2, pp. 52–205.
13. A. K. Sharma *et al.*, *Bird Conserv. Int.* **24**, 282 (2014).
14. S. A. Kools *et al.*, *Integr. Environ. Assess. Manag.* **4**, 399 (2008).
15. Committee for Medicinal Products for Veterinary Use, "Revised guideline on environmental impact assessment for veterinary medicinal products in support of the VICH guidelines GL6 and GL38" (EMEA/CVMP/418282/2005-Rev. 1, 2008; www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/10/WC500004389.pdf.
16. V. Naidoo *et al.*, *Biol. Lett.* **6**, 339 (2010).
17. I. Zorrilla *et al.*, *Conserv. Biol.* (2014). 10.1111/cobi.12417
18. S. A. Beynon, *Vet. Parasitol.* **189**, 125 (2012).
19. F. M. Aarestrup, H. C. Wegener, P. Collignon, *Expert Rev. Anti Infect. Ther.* **6**, 733 (2008).
20. A. Y. C. Tong *et al.*, *Environ. Int.* **37**, 292 (2011).
21. M. Ågerstrand *et al.*, *Environ. Int.* **35**, 778 (2009).
22. C. G. Daughton, *Environ. Health Perspect.* **111**, 757 (2003).
23. M. Baron, *Waste Biomass Valor.* **3**, 395 (2012).
24. C. G. Daughton, *Environ. Health Perspect.* **111**, 775 (2003).

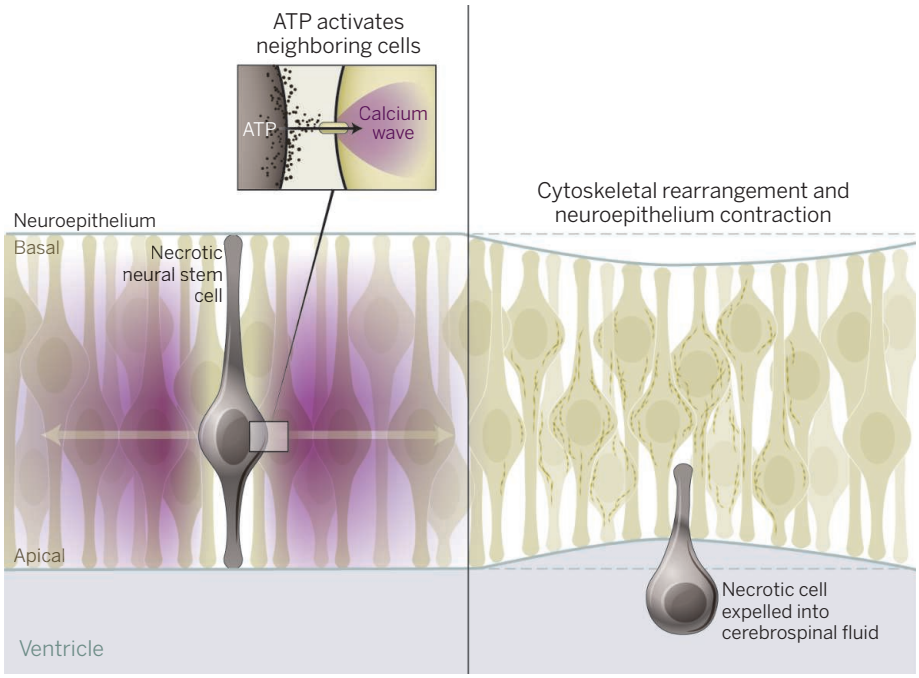
ACKNOWLEDGMENTS

We thank C. Meteyer and M. Lindsjö for manuscript assistance. The views expressed in this article are those of the authors and do not necessarily represent those of their institutions. A.M. was supported by a Ramón y Cajal research contract from the Ministerio de Economía y Competitividad (RYC-2012-11867). J.A.D. and J.A.S.-Z. were supported by projects CGL2012-40013-C02-01-02 and RNM1925.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/346/6215/1296/suppl/DC1

Published online 4 December 2014
10.1126/science.1260260



Ejected. A necrotic neural stem cell in the neuroepithelium of a developing vertebrate brain releases ATP. This activates purinergic receptors on neighboring stem cells, which triggers a calcium wave. Subsequent cytoskeletal rearrangement causes the neuroepithelium to contract, ejecting the necrotic cell.

NEURODEVELOPMENT

Tossed out to save the masses

Dead stem cells are expelled from the vertebrate brain

By **Federico Calegari**

During the 6th century, the Syriac historian John of Ephesus described how the citizens of Constantinople disposed of the corpses of the deceased inhabitants after the outbreak of a pandemic spreading through the Byzantine and Sassanid empires. Unable to cope with the numbers, all mass graves overflowing, the emperor Justinian ordered the dead to be thrown into the waters of the northern shore of the Golden Horn (*1*). Fifteen centuries later, Herrgen *et al.* (*2*) describe a similar behavior by neural stem cells of the developing vertebrate brain: burying their dead at sea by throwing them into the ventricular cerebrospinal fluid.

During development of the central nervous system in all vertebrates, neural stem cells form a densely packed and highly elongated sheet of tissue called the neuroepithelium (*3*). Within this tissue, the nuclei of neural stem cells continuously move from the boundary with the ventricle (apical) to the pial surface (basal) and then back again, completing a round of migration at every cell division cycle (*4*, *5*). This dynamic arrangement of cells (cell motion and shape) has long been proposed to represent a key step toward increasing brain size during evolution. It optimizes the number of neural stem cells that can fit within an apical unit area of tissue (*6*, *7*).

In such crowded whereabouts, within the midbrain of a developing tadpole (*Xenopus laevis*), Herrgen *et al.* killed single neural stem cells. This was accomplished by inserting a glass micropipette through the apical, ventricular-facing side of the neuroepithelium. This cell damage coincided with the onset of a calcium wave—an increase in intracellular calcium that propagated in a wave-like fashion across many cells. The wave was of reproducible origin, speed, and duration and was specifically caused by the death of the pierced cell; when a target cell was not killed and the micropipette merely contacted the cell, no calcium wave was generated. Moreover, the correlation between cell death and calcium waves was faithfully reproduced by more sophisticated means of killing that did not require physical contact, such as electric shocks or laser irradiation.

DFG–Research Center and Cluster of Excellence for Regenerative Therapies, Technische Universität Dresden, 01307 Dresden, Germany. E-mail: federico.calegari@crt-dresden.de

ILLUSTRATION: V. ALTOUNIAN/*SCIENCE*

Published by AAAS

When the calcium wave reached its peak of lateral diffusion of about 170 μm in 20 s (involving about 50 neural stem cells), the whole neuroepithelium started to contract by about 10% along its apical-basal axis—that is, along the same route followed by the nuclei of neural stem cells during their cyclic migration during cell cycle progression. These contractions resulted in the expulsion of the single dead cell at the apical side of the neuroepithelium, as if the remaining neural stem cells wanted to get rid of the remains of their unfortunate fellow cell by throwing it into the cerebrospinal fluid. It may be that the apical side is more convenient for this release because a dying cell loses its apically located cell-cell junctions (important for epithelial integrity) and because the basal domain of the neuroepithelium is delimited by a stiff pial membrane.

...the remaining neural stem cells wanted to get rid of ... their unfortunate fellow cell by throwing it into the cerebrospinal fluid.

Calcium waves play a key role in wound repair. Herrgen *et al.* observed that the calcium wave elicited at the onset of cell death was triggered by the release of adenosine triphosphate (ATP) from the necrotic cell itself. Moreover, the neuroepithelial contractions were autonomous because brain explants in culture contracted equally well. The authors determined that the contractility was mediated by a reorganization of cytoskeletal components in the neural stem cells surrounding the necrotic cell (see the figure). These components are also important in coordinating the cyclic migration of neural stem cell nuclei during cell cycle progression (*5*).

Is the motivation for the neuroepithelium to eject a necrotic cell quickly (within a minute) similar to that of the Constantinople citizens who threw their dead into the sea? To address this, Herrgen *et al.* induced cell death while preventing contractions—and hence cell expulsion—through the use of drugs that antagonize cytoskeletal reorganization. They observed that this inhibition correlated with an increase in secondary cell death in the area surrounding the necrotic, but unexpelled, cell.

Admittedly, calcium waves, tissue contractions, and cell expulsion have been described before (*8*–*10*). Yet this study elegantly links these events in a functional context that is relevant in tissue protection upon injury. Indeed, Herrgen *et al.* propose that cell contraction and expulsion constitute a mechanism by which the tissue protects itself from the toxic substances released by the necrotic cell. Their results with amphibians are consistent with previous observations in various epithelia of invertebrates (*10*).

In the *Xenopus* tadpole, the thickness of the neuroepithelium is in the range of 30 to 50 μm; this makes it easy to imagine how a contraction may result in the expulsion of a necrotic cell whose volume is comparable to that of a spheroid 5 to 10 μm in diameter. It is slightly more difficult to imagine this process occurring in other species, notably among mammals, in which a similarly sized spheroid would need to be expelled from a neuroepithelium that may be several times as thick at a comparable developmental stage. Future studies in different species and/or tissues are needed to determine how widespread this mechanism is in the animal kingdom.

An additional intriguing question arising from this study is why cells dying through programmed cell death (apoptosis) did not trigger the same calcium waves, contractions, and expulsions that were generated by cells dying through necrosis. Clearly, necrotic cell death due to injury is fundamentally different from physiological apoptosis, the latter being comparable to a carefully planned suicide sporadically occurring in all tissues, including the neuroepithelium (*11*). Yet all potentially toxic components of a dying cell will be released sooner or later within the tissue. As a result, it seems reasonable to conclude that cells should try to get rid of the remains of a dead cell, no matter the cause of death. One possible exception to this would involve a carefully planned suicide, allowing the cell enough time to eliminate toxic molecules and make sure that no other cell is harmed as collateral damage. If this is the case, then understanding the mechanisms that "make sure that no one else is hurt" may be extremely important for developing tissue-protective approaches in many paradigms of injury or disease. With their study, Herrgen *et al.* may have provided a fundamental insight toward reaching this goal. ■

REFERENCES

1. J. Ephesus , Musei Vaticani, fragment IIE (circa 542 CE).
2. L. Herrgen *et al.*, *Dev. Cell* 10.1016/j.devcel.2014.10.012 (2014).
3. E. Taverna, M. Götz, W. B. Huttner, *Annu. Rev. Cell Dev. Biol.* **30**, 465 (2014).
4. F. C. Sauer, *J. Comp. Neurol.* **62**, 377 (1935).
5. H. O. Lee, C. Norden, *Trends Cell Biol.* **23**, 141 (2013).
6. I. H. Smart, *J. Anat.* **116**, 67 (1973).
7. J. H. Lui, D. V. Hansen, A. R. Kriegstein, *Cell* **146**, 18 (2011).
8. D. F. Owens, A. R. Kriegstein, *J. Neurosci.* **18**, 5374 (1998).
9. N. Gorfinkiel, G. B. Blanchard, *Curr. Opin. Cell Biol.* **23**, 531 (2011).
10. Y. Gu, J. Rosenblatt, *Curr. Opin. Cell Biol.* **24**, 865 (2012).
11. D. Thomaidou *et al.*, *J. Neurosci.* **17**, 1075 (1997).

10.1126/science.aaa3261

GENETICS

The battle for iron

Evolutionary analysis shows how a mammalian iron transport protein evolves to avoid capture by bacteria

By **Andrew E. Armitage** *and* **Hal Drakesmith**

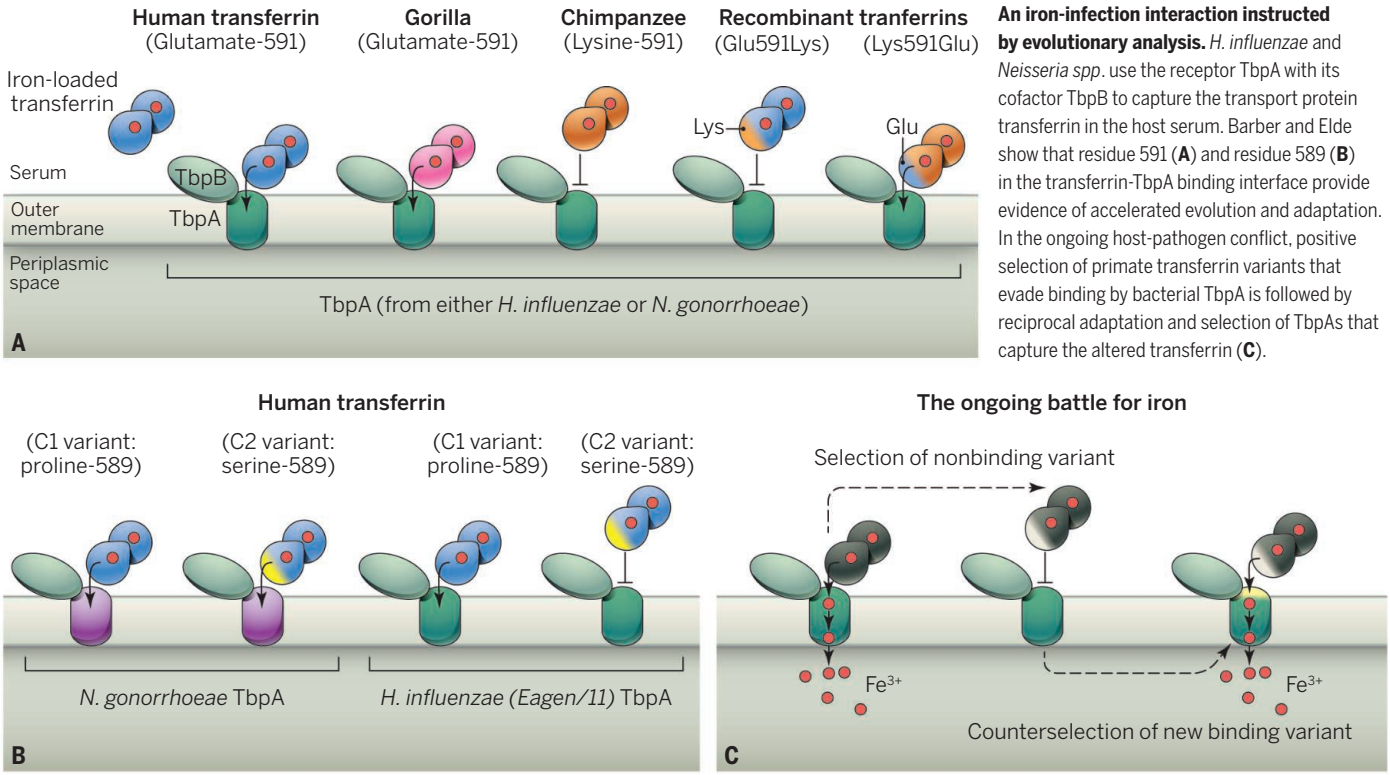
Infections exert selection pressures on host and pathogen alike as they both struggle to survive. Host genes important for the immune response to infection often show enhanced rates of diversification, termed positive selection, between and even within species (*1*, *2*). Sequence analysis can reveal "hotspots" in proteins that have been selected by evolution and are critical to repel infections. For example, proteins that recognize viruses exhibit amino acid diversity in the specific regions that directly contact components of the virus (*1*). Of course, viruses can and do counteradapt through selection of variants that evade recognition by defense proteins (*1*). Bacterial infections are also likely to positively select mammalian genes important for immune defense. On page 1362 of this issue, Barber and Elde (*3*) provide one striking example of this that is centered on iron.

Why iron? Humans need it to grow, but so do almost all organisms that infect us, because it supports basic cellular processes such as electron transport and DNA metabolism. The virulence of infectious organisms is often highly dependent on their ability to assimilate host iron (*4*). However, the host fights back: Upon detection of infection, the innate immune system interfaces with the circuitry that controls iron transport. Increased synthesis of several iron- and heme-binding proteins and of the iron regulatory hormone hepcidin (*5*) lowers serum iron availability for blood-borne pathogens. Another key component of mammalian iron transport is the serum protein transferrin, and it is this protein that Barber and Elde focus on.

Transferrin has two lobes, N and C, each of which binds one ferric ion (see the figure). Transferrin receptors at the cell surface enable cellular iron uptake. The extremely high affinity of transferrin for iron efficiently withholds iron from many pathogens. Indeed, when Schade and Caroline reported the purification of transferrin in 1946 they

Downloaded from www.sciencemag.org on December 14, 2014

Published by AAAS



An iron-infection interaction instructed by evolutionary analysis. *H. influenzae* and *Neisseria spp*. use the receptor TbpA with its cofactor TbpB to capture the transport protein transferrin in the host serum. Barber and Elde show that residue 591 (**A**) and residue 589 (**B**) in the transferrin-TbpA binding interface provide evidence of accelerated evolution and adaptation. In the ongoing host-pathogen conflict, positive selection of primate transferrin variants that evade binding by bacterial TbpA is followed by reciprocal adaptation and selection of TbpAs that capture the altered transferrin (**C**).

noted its likely relevance for delivering iron around the body and, prophetically, "the bacteriostatic action it exerts upon iron-sensitive pathogens" (*6*). However, many bacteria, such as *Haemophilus influenzae* and *Neisseria* species, have a successful countermeasure: They encode their own transferrin capture machinery, notably TbpA (transferrin-binding protein A), which seizes transferrin from serum.

To assess transferrin variation through evolutionary time, Barber and Elde sequenced transferrin orthologs from 21 species, including humans, gorillas, chimpanzees, and Old and New World monkeys. (Orthologs are genes in different species that evolved from a common ancestral gene by speciation.) They then calculated signatures of positive selection and combined this information with crystal structures of transferrin bound to TbpA (*7*). The authors identified 18 rapidly evolving sites in transferrin; 16 of these sites are located in the C lobe that binds to TbpA. Only one site contacts the mammalian transferrin receptor, whereas 14 of the 16 C-lobe sites could be directly mapped to the interface with *N. meningitidis* TbpA. Thus, evolution of transferrin has likely been driven by its interaction with TbpA and ancestral TbpA variants.

To further assess the consequences of transferrin sequence variation, Barber and Elde generated recombinant primate transferrins and tested their interactions with TbpA from *N. gonorrhoeae* and *H. influenzae*. Both TbpAs bound transferrin from humans and gorillas but not the closely related chimpanzee transferrin. The amino acid at position 591 of transferrin is different between humans and chimpanzees; swapping this single amino acid between the different orthologs was sufficient to inhibit binding of human transferrin to *N. gonorrhoeae* and *H. influenzae* TbpA and to restore binding of chimpanzee transferrin (see the figure, panel A).

The two most common variants of human transferrin, C1 and C2, differ at only one amino acid, position 589, which like position 591 lies at the binding interface. When the authors assessed C1 and C2 variants for binding TbpA from different bacteria, they found that all TbpAs bound C1 transferrin; in contrast, C2 transferrin did not bind TbpAs from two *H. influenzae* strains (see the figure, panel B). Thus, a hotspot at positions 589 and 591 in the large (2500 Å^2) TbpA-transferrin interface has a powerful effect on the interaction between these two proteins.

The authors also investigated TbpA sequence variants for indications of positive selection. Rapidly evolving sites in both *Neisseria* and *H. influenzae* TbpA sequences overlapped with the regions of TbpA known to bind transferrin. Moreover, replacement of a *N. gonorrhoeae* TbpA amino acid under selection with the corresponding amino acid from *H. influenzae* reduced binding to human transferrin. These findings suggest that the TbpA-transferrin binding interface is the subject of an evolutionary arms race, in which host transferrin variants are selected to avoid bacterial recognition, only to be faced with counterselected TbpA adapted to recapture the altered transferrin (see the figure, panel C).

The study by Barber and Elde shows that resisting bacterial capture of transferrin is important enough to leave a clear evolutionary imprint on primate and hominoid genomes. Further research should investigate whether humans carrying particular transferrin variants that are differentially susceptible to specific TbpA-encoding bacteria. Other components of mammalian iron homeostasis may also be shaped by selection pressure from infectious disease. Furthermore, in the era of increasing antibiotic resistance, targeting microbial iron acquisition, guided by positive selection analysis and structural biology, may aid the development of new antimicrobial agents. ■

MRC Human Immunology Unit, Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, OX3 9DS, UK. E-mail: alexander.drakesmith@ndm.ox.ac.uk

REFERENCES

1. M. D. Daugherty, H. S. Malik, *Annu. Rev. Genet.* **46**, 677 (2012).
2. E. C. Holmes, *PLOS Biol.* **2**, e307 (2004).
3. M. F. Barber, N. C. Elde, *Science* **346**, 1362 (2014).
4. E. Carniel, *Int. Microbiol.* **2**, 161 (1999).
5. H. Drakesmith, A. M. Prentice, *Science* **338**, 768 (2012).
6. A. L. Schade, L. Caroline, *Science* **104**, 340 (1946).
7. N. Noinaj *et al*., *Nature* **483**, 53 (2012).

10.1126/science.aaa2468

ILLUSTRATION: P. HUEY/*SCIENCE*

Published by AAAS

BOOKS *et al.*

GLOBAL HEALTH

To run where the brave dare not go

By **Kai Chen**

Over the past several decades, the international humanitarian aid organization Médecins Sans Frontières (Doctors Without Borders, hereinafter MSF) has been widely discussed in books on public health and international relations. In *Doctors Without Borders: Humanitarian Quests, Impossible Dreams of Médecins Sans Frontières,* Renée C. Fox forgoes the traditional narrative of these books, which tend to focus on individual physicians and patients, in favor of discussing the most distinctive features of its organizational culture. She supplements the prose with a collection of photographs and illustrations.

Fox, a medical sociologist by training, has conducted participant observation within MSF for almost 20 years. From an insider-outsider perspective, she stresses that the core principles of MSF—medical ethics, impartiality and neutrality, independence, bearing witness, and internal accountability—have been internalized into the organizational culture.

She begins by analyzing the firsthand accounts (e.g., field blogs and notes) of MSF staff. In addition to being a powerful way to bear witness, the field blogs reveal a commitment to continuous reevaluation of what motivates staff to serve in MSF as they grapple with limitations that impair their ability to relieve suffering and save lives. "We do everything possible for each individual patient," reads one blog post. "I had to hold back tears, though, when I realized that 'everything possible' is likely not going to be enough in the long run."

In the following pages, Fox chronicles the salient events of a number of debates that have occurred within MSF in recent years, which she frames as reflections of the organization's commitment to internal accountability and nonhierarchy. One of these debates occurred at the "La Mancha" conference in Luxembourg in 2006, so called because of the region's association with the fictional adventurer Don Quixote, whose idealism led to "misperceptions and imperfections"—a theme that resonated with the conference's goal of critical self-evaluation. The debate focused on the underrepresentation of "national staff"—those working on MSF projects in their native countries—within the organization's leadership. A contingent of MSF members argued that the creation of more non-Western branches would help the organization realize medicine "without borders." As a result of the discussions at La Mancha, the International Council established MSF South Africa as a delegate office in 2009, and discussions are under way regarding new branch offices in mainland China.



MSF's commitment to participatory democracy has at times impaired decision-making within the organization, a phenomenon that was satirized at the meeting by former MSF Head of Mission Samuel Hanryon. Fox describes a cartoon drawn by Hanryon, which depicts a group of MSF members in a fistfight. The overhead caption reads: "Do we agree to disagree publicly? Yes, we agree!"

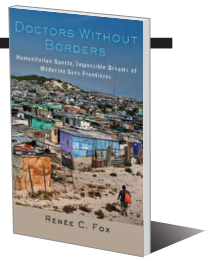
Next, Fox examines MSF's responses to the epidemics of HIV/AIDS in South Africa and tuberculosis in Russia and the role that partnerships (or the lack thereof) played in each case. Until the late 20th century, MSF primarily operated as an independent organization, but it became clear that partnerships with both public and private entities would become necessary if the organization hoped to tackle chronic epidemics. Fox describes an example of a successful collaboration, known as the MSF Khayelitsha program. Established in April 2000, MSF Khayelitsha consists of HIV/AIDS clinics in three community health centers in South Africa. This program was made possible by partnerships with national and local governments and the Brazilian pharmaceutical company Farmanguinhos. In contrast, without the endorsement of the Russian Ministry of Health, MSF was ultimately forced to terminate a program for Russian prisoners with multidrug-resistant tuberculosis.

Not all partnerships are embraced by MSF, however. Citing incidents in Kosovo, Afghanistan, and Iraq, Fox discusses the controversy surrounding military-humanitarian coalitions. Not surprisingly, most MSF members consider military partnerships to be "incompatible with true humanitarian ideals" and fear that they jeopardize the MSF commitment to independence from political influence.

Though not covered in the book, MSF staff are actively engaged in combating the current Ebola epidemic in West Africa, having cared for more than two-thirds of the reported cases in this region (*1*). This situation could require the organization to achieve a timely consensus on how to reallocate resources from longer-term initiatives. Although the provisional 2014 budget for MSF's Ebola response in West Africa has reached €51 million (*2*), more aid funding could be needed, as it is estimated that between 550,000 and 1.4 million people in Africa could be infected by January 2015 (*3*).

Carefully researched and delightfully written, *Doctors Without Borders* establishes a new bar for those who would cover Médecins Sans Frontières in the future. This book will take its due place as one of the most comprehensive works on MSF.

Doctors Without Borders Humanitarian Quests, Impossible Dreams of Médecins Sans Frontières
Renée C. Fox
Johns Hopkins University Press, 2014. 326 pp.



REFERENCES

1. M. Underwood, The perfect storm: How Ebola has taken hold of West Africa, *The National*, 25 September 2014).
2. Médecins Sans Frontières, Ebola Crisis update 30th October 2014; www.msf.org/article/ebola-crisis-update-30th-october-2014.
3. G. Dyer, First US Ebola case diagnosed in Texas, *Financial Times*, 1 October 2014.

The reviewer is at the College of Public Administration, Zhejiang University, Hangzhou, Zhejiang 310058, China. E-mail: chenkai@zju.edu.cn

10.1126/science.aaa1486

PHOTO: ISOPIX/REX FEATURES/AP PHOTOS

Downloaded from www.sciencemag.org on December 14, 2014

Published by AAAS



EVOLUTIONARY BIOLOGY

The animal arms race

By **Viviane Callier**

The visually striking weapons of living creatures—antlers, horns, teeth, and spines, to name a few—have fascinated biologists for a long time. These are most often used in the context of competition for resources or access to females. But why is it that some animals do not have weapons, or have only modest ones, whereas others have developed massive and evolutionarily expensive weaponry? In *Animal Weapons: The Evolution of Battle*, evolutionary biologist Douglas Emlen investigates the specific conditions under which extreme animal weapons evolve. A skilled storyteller, Emlen takes the reader to Panama to observe dueling Harlequin beetles; to Tanzania to collect elephant excrement—and the dung beetles that live on it; and to tropical forests where stalk-eyed flies defend their harems of females. The weapons discussed in these stories are illustrated in stunning color plates.

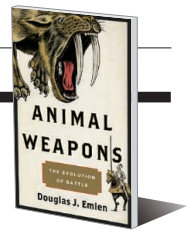
Emlen argues that three preconditions are necessary for triggering an arms race that leads to extreme animal weapons. First, there must be competition for a limited resource, such as food or access to reproductive females. Second, the limited resource must be confined to well-defined locations so that one animal can control access to it (i.e., the resource must be "defensible"). Third, males must fight one-on-one; guerilla-style warfare does not lead to the evolution of extreme weapons.

Emlen's third condition is inspired by a set of mathematical models of human warfare published by the British engineer and mathematician Frederick Lanchester in 1956 (*1*). Before the invention of firearms, soldiers in battle fought one-on-one, and soldiers were limited to attacking one opponent at a time. The winning army was determined by the effectiveness of each soldier, multiplied by the number of soldiers. In this context, better weapons (and better protective armor) made a significant difference to the outcomes of a battle. On the other hand, the invention of firearms was a game changer because it allowed a single soldier to attack multiple foes simultaneously. According to Lanchester's model, the effectiveness of an army that uses firearms is proportional to the number of soldiers squared. In this context, the number of soldiers matters more than the relative quality of the firearm. Weapons (and armor) matter most in one-on-one battles. Emlen uses this rationale to explain why species that go head-to-head are more likely to develop big weapons.

Emlen's thesis not only makes sense of the biological diversity of animal weapons—it also helps to explain human political and military dynamics, which he discusses in the last quarter of the book. Indeed, escalation is not just a phenomenon of biological evolution but also a feature of human culture and politics. The book's treatment of military escalation helps make sense of the post–Cold War proliferation of acts of terrorism and the development of new and more deadly weapons. According to Emlen, terrorism and weapons of mass destruction are also game changers that circumvent the traditional models of warfare. These new forms of battle therefore require a different suite of weapons and defenses.

We may not face another nuclear escalation, but the development of new strategies and weapons will surely continue. It is difficult to estimate how long it will take before these new tactics and weaponry raise the stakes of war even further, but as Emlen concludes, "We're not likely to survive another arms race."

The reviewer is at The Ronin Institute for Independent Scholarship, Montclair, NJ 07043, USA.
E-mail: vcallier@gmail.com



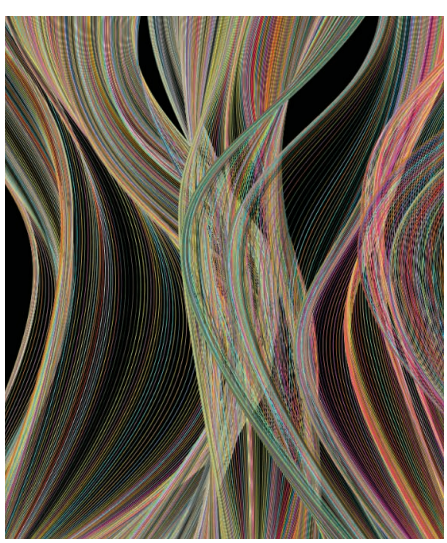
Animal Weapons
The Evolution of Battle
Douglas J. Emlen
Henry Holt, 2014. 288 pp.

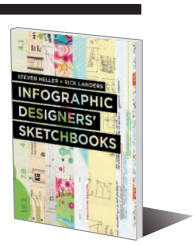
REFERENCES

1. F. W. Lanchester, in *The World of Mathematics*, vol. 4, J. R. Newman, Ed. (Simon and Schuster, New York, 1956), pp. 2138–2157.

10.1126/science.1260024

BOOKS IN BRIEF





Infographic Designers' Sketchbooks
By Steven Heller and Rick Landers
Princeton Architectural Press, 2014, 351 pp.

Featuring the work of more than 70 graphic designers and illustrators from around the world, this book offers a behind-the-scenes look at the process of developing concise and accurate infographics. Rendered in large format, the colorful doodles, drafts, and mock-ups are accompanied by short descriptions of the designers and their methods. The result is a collection that you may find yourself turning to for inspiration the next time you are faced with distilling complex data into a clear and engaging figure.

10.1126/science.aaa3324

PHOTO: (TOP) IGOR SIWANOWICZ; IMAGE (RIGHT) CREDIT: ALEKSANDAR MAČAŠEV

Downloaded from www.sciencemag.org on December 14, 2014

Published by AAAS

LETTERS

Edited by **Jennifer Sills**

Recognizing the risks of brain stimulation

IN THE REVIEW "Plasticity of the aging brain: New directions in cognitive neuroscience" (31 October, p. 579), A. Gutchess states that both repetitive transcranial magnetic stimulation (rTMS) and transcranial direct current stimulation (tDCS) are safe for human participants when operated according to the guidelines. With respect to the aging brain, it is simply too premature to draw this conclusion, given the lack of studies on the safety of neurostimulation in this segment of the population. Serious side effects of rTMS in young to middle-aged participants have been reported (*1*, *2*), and milder side effects have been reported in both rTMS and tDCS (*1*–*4*). Recently, Davis (*1*) pledged extreme caution in using brain stimulation because of its unknown (side) effects.

Neurostimulation on senior participants could be even more dangerous, given that the aging brain is known to be more vulnerable than the young brain (*5*). Although we see the potential of brain stimulation in the aging brain, we are deeply concerned that Gutchess seems to severely underestimate the risks. We strongly recommend that the safety issues are clarified first—including the formation of strict guidelines on brain stimulation in senior participants to ensure their safety—before brain stimulation is used as an intervention technique in the aging brain.

Maurits van den Noort,[1]* Sabina Lim,[1] Peggy Bosch[2]

[1]Research Group of Pain and Neuroscience, Kyung Hee University Seoul 130-701, Republic of Korea. [2]Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, 6525, HR Nijmegen, Netherlands.

*Corresponding author. E-mail: info@mauritsvandennoort.com

REFERENCES

1. N. J. Davis, *Front. Hum. Neurosci.* **8**, 600 (2014).
2. S. Rossi, M. Hallett, P. M. Rossini, A. Pascual-Leone, *Clin. Neurophysiol.* **120**, 2008 (2009).
3. C. Poreisz, K. Boros, A. Antal, W. Paulus, *Brain Res. Bull.* **72**, 208 (2007).
4. M. A. Nitsche *et al.*, *Brain Stimul.* **1**, 206 (2008).
5. E. Sibille, *Dialogues Clin. Neurosci.* **15**, 53 (2013).

Response

IN THEIR LETTER, van den Noort *et al.* highlight the importance of considering safety issues in research and clinical interventions that use neurostimulation, particularly with older adults. Several of the papers cited in my Review did note the relative safety of these methods (*1*), even in regard to the study of older adults (*2*, *3*), but the caveat "when operated according to the guidelines" is an important point that should not be overlooked. Although it was outside the scope of my Review to examine safety issues in depth, I encourage interested readers to consult the primary sources for systematic study of these issues (*4*–*6*).

Some studies of outcomes (*7*) have considered a broad age range of adults and, as I reviewed, research studies are beginning to investigate older adults. However, I agree with van den Noort *et al.* that age should be systematically considered in evaluations of safety concerns, just as Davis (*8*) underscores the need for more translational studies from adults to children in his plea for caution.



Repetitive transcranial magnetic stimulation.

van den Noort *et al.* are right to point out that low risk is not the same thing as no risk, and thus it is crucial for researchers and clinicians to take steps to minimize risks (*4*). As with any research protocol or medical treatment, potential participants and patients should consult with experts about their individual circumstances. Researchers and clinicians should screen patients with care and fully inform them of potential risks and side effects before they consent to a procedure.

Angela Gutchess

Department of Psychology and Volen National Center for Complex Systems, Brandeis University, Waltham, MA 02453, USA and Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Boston, MA 02114, USA. E-mail: gutchess@brandeis.edu

REFERENCES

1. B. A. Coffman, V. P. Clark, R. Parasuraman, *Neuroimage* **85**, 895 (2014).
2. A. Pascual-Leone *et al.*, *Brain Topogr.* **24**, 302 (2011).
3. C. Freitas, F. Farzan, A. Pascual-Leone, *Front. Neurosci.* **7**, 42 (2013).
4. S. Rossi, M. Hallett, P. M. Rossini, A. Pascual-Leone, *Clin. Neurophysiol.* **120**, 2008 (2009).
5. E. M. Wassermannn, *Electroencephalogr. Clin. Neurophysiol.* **108**, 1 (1998).
6. C. Poreisz, K. Boros, A. Antal, W. Paulus, *Brain Res. Bull.* **72**, 208 (2007).
7. J. P. O'Reardon *et al.*, *Biol. Psychiat.* **62**, 1208 (2007).
8. N. J. Davis, *Front. Hum. Neurosci.* **8**, 600 (2014).

Reproducibility in ecological research

THE EDITORIAL BY M. McNutt ("Journals unite for reproducibility," 7 November, p. 679, published online 5 November) describes an updated version of the solution from journals, including *Science* and *Nature*, for reproducibility in biomedical research. If the new policy is to be widely implemented by scientific journals, then the changes must be consistent and mandatory.

Reproducibility is not just relevant for biomedical research. Ecology and biodiversity scientists are also increasingly concerned about issues of reproducibility and data sharing (*1*–*3*). Reproducibility is challenging for ecology and biodiversity science because experiments are usually undertaken under natural conditions in the field, across different spatial and temporal scales (*4*). For this reason, there are no prevalent recommendations for quality control and few for sharing data in these disciplines [one data-sharing option is Dryad (*5*)]. It is time for ecologists and biodiversity scientists to work out some principles and guidelines equivalent to that in biomedical research; scientists and journals can start with a stringent data-sharing policy.

Xiaolei Huang

College of Plant Protection, Fujian Agriculture and Forestry University, Cangshan District, Fuzhou 350002, China and Key Laboratory of Integrated Pest Management for Fujian-Taiwan Crops, Ministry of Agriculture, Fuzhou 350002, China. E-mail: huangxl@fafu.edu.cn

REFERENCES

1. O. J. Reichman, M. B. Jones, M. P. Schildhauer, *Science* **331**, 703 (2011).
2. X. Huang *et al.*, *Conserv. Lett.* **5**, 399 (2012).
3. P. A. Soranno, K. S. Cheruvelil, K. C. Elliott, G. M. Montgomery, *BioScience* 10.1093/biosci/biu169 (2014).
4. M. J. Ryan, *Science* **334**, 1229 (2011).
5. Dryad (http://datadryad.org).

PHOTO: RICHARD T. NOWITZ / SCIENCE SOURCE

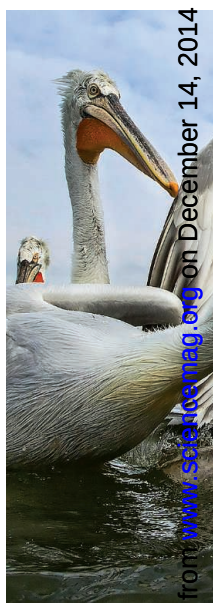
Downloaded from www.sciencemag.org on December 14, 2014

Published by AAAS









Downloaded from www.sciencemag.org on December 14, 2014

A flock of Genomes

by **Guojie Zhang, Erich D. Jarvis,** *and* **M. Thomas P. Gilbert**

Published by AAAS









PHOTO: (CLOCKWISE) © PAUL SOUDERS/CORBIS; © PETER JOHNSON/CORBIS; © ALFRED SCHAUHUBER/IMAGEBROKER/CORBIS; © MICHAEL MELFORD/NATIONAL GEOGRAPHIC SOCIETY/CORBIS; © JAN VAN DER GREEF/BUITEN-BEELD/MINDEN PICTURES/CORBIS; © JARED HOBBS/ALL CANADA PHOTOS/CORBIS

INSIDE

OPINION

Valuing collections *p. 1310*

RESEARCH ARTICLES

Comparative genomics reveals insights into avian genome evolution and adaptation *p. 1311*

Whole-genome analyses resolve early branches in the tree of life of modern birds *p. 1320*

RESEARCH ARTICLE SUMMARIES

Complex evolutionary trajectories of sex chromosomes across bird taxa *p. 1332*

Convergent transcriptional specializations in the brains of humans and song-learning birds *p. 1333*

Core and region-enriched networks of behaviorally regulated genes and the singing genome *p. 1334*

Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs *p. 1335*

Evidence for a single loss of mineralized teeth in the common avian ancestor *p. 1336*

Statistical binning enables an accurate coalescent-based estimation of the avian tree *p. 1337*

SEE ALSO

▸ NEWS STORY P. 1275
▸ REVIEW p. 1341
▸ PODCAST

Characterization of genomic biodiversity through comprehensive species sampling has the potential to change our understanding of evolution. To study evolution across a major vertebrate class, dissect the genomics of complex traits, and resolve a centuries-old debate on the avian species tree, we formed a consortium focused on the sequencing and analyses of at least one genome per avian order. The resulting data set of 48 consistently annotated bird genomes spans 32 of the 35 recently proposed avian orders,* including all 30 neognath orders, and thus represents a wide range of avian evolutionary diversity. Our consortium's analyses have resulted in eight papers published today in *Science*, as well as 20 papers in other journals [avian.genomics.cn/en]. These include two flagship papers: one exploiting genomic-scale data to generate a highly supported avian order phylogeny that resolves many debates on the timing and topology of their radiation; the other a comparative genomic analysis exploring avian genome evolution and the genetic basis of complex traits. Other studies in *Science* describe convergent brain regions and gene expression for avian song learning and human speech, the singing activated genome in songbirds, complex evolutionary trajectories of avian sex chromosomes, a single loss of teeth in the ancestor of modern birds, the genomes of their closest extant outgroup (crocodilians) and inferred dinosaur ancestor, and computational methods developed for large-scale genomic analyses. Studies in companion papers explore the genomic adaptations of penguins, genomics of nearly extinct species, lineage-specific selection in birds, paleoviral infiltration in bird genomes, and many other questions. Thus, this study of a major vertebrate class highlights the future promise of large-scale comparative genomics, and we hope sets the stage for an approach for sequencing and analyses of many more genomes of birds and other vertebrate lineages.

*J. Cracraft, in *The Howard and Moore Complete Checklist of the Birds of the World*, E. C. Dickinson, J. C. J. Remsen, Eds. (Aves Press, Eastbourne, UK, 2013), pp. xxi–xliii.

Published by AAAS

OPINION

Valuing collections

This year brought dismal news about the world's birds: They are vanishing at an alarming rate. Across 25 European countries, about 420 million fewer birds are present today than in 1980, a 20% decrease, especially in the 36 most common species. In North America, *The State of the Birds Report 2014* indicates that over the past 40 years, the numbers of individuals across 33 species are also down by hundreds of millions. Such assessments highlight the urgency of determining the precise causes of these declines. The knowledge gleaned from the Avian Phylogenomics Project, coupled with ecological and population analyses, should provide new insights into the factors that influence bird declines and extinctions. As the project progresses over the next few years, over 60% of tissue samples for the avian analyses will be derived from archived museum collections. In this era of deteriorating natural environments, a pressing challenge is to continue to build scientific collections for future needs.



"...the most pressing challenge is to build collections for future needs..."

Museum collections, and the species they represent, provide windows into the past, inform about the present, and help predict the future of natural habitats and human-altered environments. They are the common language of the biological sciences. An antiquated view of collections suggests drawers of bird skins, empty shells, and dried plants. However, current collections also include living specimens, spirit-preserved samples, deep-frozen tissues, and DNA. These irreplaceable biomaterials are invaluable representatives of Earth's biodiversity, and together with their associated metadata are archived ex situ for long-term documentation, public education and exhibition, and scientific and applied research. Although the exact number of collections maintained in museums, botanic gardens, and universities is unknown, estimates as high as three billion specimens suggest the magnitude of this storehouse of information about the natural world.

Many scientists continue to use collections to discover, describe, and document plants and animals with time-proven methods. At the same time, the application of new and multiple technologies to study specimens is blossoming. For example, much of our current understanding of some recently extinct species, such as the Tasmanian tiger, the Caribbean monk seal, and the passenger pigeon, has directly resulted from genomic information extracted from museum collections. And combining DNA-, amino acid– and isotope-based analyses of a few grams of bone from a historical specimen of an endangered Pacific seabird, the Hawaiian petrel, has illuminated aspects of the bird's diet, past population demographics, food chain dynamics, and the deleterious impacts of industrial fishing on this oceanic predator. Museums are becoming an unparalleled resource of tissue samples for large-scale genomic studies of animals and plants.

Yet contributions to genomic studies are but one use of museum collections. Extinctions due to human impacts are also readily studied through historical records preserved in scientific collections. These records reveal former patterns of geographic distributions and population abundances of species that today are threatened or extinct. Museum collections equally contribute to the discovery of previously unknown species, such as the olinguito, a carnivorous mammal from South America.

Most museum specimens were not collected for the purposes for which they are now used. Innovation in technologies will continue to reveal new information previously unanticipated in scientific specimens. Therefore, the most pressing challenge is to build collections for future needs that maximize access and benefit-sharing for all. Collections must be sustained for the long term, which will require increased funding for their physical and scientific curation. A formidable task that has only just begun is the computerized inventory and digitization of the wealth of information that collections represent. At present, most of these data are not accessible electronically or online.

The worldwide decline of birds is just one part of a large and grim picture; the number of individuals of all wild animals on Earth has decreased by 50% since the 1970s. This drastic decline underscores the vital inherent value of museum collections today, tomorrow, and into the future.

– W. John Kress



W. John Kress is the Interim Under Secretary for Science at the Smithsonian Institution and Distinguished Scientist and Curator of Botany at the National Museum of Natural History, Smithsonian Institution, Washington, DC. E-mail: kressj@si.edu

PHOTOS: (INSET) COURTESY OF THE SMITHSONIAN INSTITUTE; (TOP RIGHT) AMANDA LUCIDON

Downloaded from www.sciencemag.org on December 14, 2014

10.1126/science.aaa4115

Published by AAAS

RESEARCH ARTICLE

Comparative genomics reveals insights into avian genome evolution and adaptation

Guojie Zhang,[1,2]*† Cai Li,[1,3]* Qiye Li,[1,3] Bo Li,[1] Denis M. Larkin,[4] Chul Lee,[5,6] Jay F. Storz,[7] Agostinho Antunes,[8,9] Matthew J. Greenwold,[10] Robert W. Meredith,[11] Anders Ödeen,[12] Jie Cui,[13,14] Qi Zhou,[15] Luohao Xu,[1,16] Hailin Pan,[1] Zongji Wang,[1,17] Lijun Jin,[1] Pei Zhang,[1] Haofu Hu,[1] Wei Yang,[1] Jiang Hu,[1] Jin Xiao,[1] Zhikai Yang,[1] Yang Liu,[1] Qiaolin Xie,[1] Hao Yu,[1] Jinmin Lian,[1] Ping Wen,[1] Fang Zhang,[1] Hui Li,[1] Yongli Zeng,[1] Zijun Xiong,[1] Shiping Liu,[1,17] Long Zhou,[1] Zhiyong Huang,[1] Na An,[1] Jie Wang,[1,18] Qiumei Zheng,[1] Yingqi Xiong,[1] Guangbiao Wang,[1] Bo Wang,[1] Jingjing Wang,[1] Yu Fan,[19] Rute R. da Fonseca,[3] Alonzo Alfaro-Núñez,[3] Mikkel Schubert,[3] Ludovic Orlando,[3] Tobias Mourier,[3] Jason T. Howard,[20] Ganeshkumar Ganapathy,[20] Andreas Pfenning,[20] Osceola Whitney,[20] Miriam V. Rivas,[20] Erina Hara,[20] Julia Smith,[20] Marta Farré,[4] Jitendra Narayan,[21] Gancho Slavov,[21] Michael N Romanov,[22] Rui Borges,[8,9] João Paulo Machado,[8,23] Imran Khan,[8,9] Mark S. Springer,[24] John Gatesy,[24] Federico G. Hoffmann,[25,26] Juan C. Opazo,[27] Olle Håstad,[28] Roger H. Sawyer,[10] Heebal Kim,[5,6,29] Kyu-Won Kim,[5] Hyeon Jeong Kim,[6] Seoae Cho,[6] Ning Li,[30] Yinhua Huang,[30,31] Michael W. Bruford,[32] Xiangjiang Zhan,[32,33] Andrew Dixon,[34] Mads F. Bertelsen,[35] Elizabeth Derryberry,[36,37] Wesley Warren,[38] Richard K Wilson,[38] Shengbin Li,[39] David A. Ray,[26]‡ Richard E. Green,[40] Stephen J. O'Brien,[41,42] Darren Griffin,[22] Warren E. Johnson,[43] David Haussler,[40] Oliver A. Ryder,[44] Eske Willerslev,[3] Gary R. Graves,[45,46] Per Alström,[47,48] Jon Fjeldså,[46] David P. Mindell,[49] Scott V. Edwards,[50] Edward L. Braun,[51] Carsten Rahbek,[46,52] David W. Burt,[53] Peter Houde,[54] Yong Zhang,[1] Huanming Yang,[1,55] Jian Wang,[1] Avian Genome Consortium,§ Erich D. Jarvis,[20]† M. Thomas P. Gilbert,[3,56]† Jun Wang[1,55,57,58,59]†

Birds are the most species-rich class of tetrapod vertebrates and have wide relevance across many research fields. We explored bird macroevolution using full genomes from 48 avian species representing all major extant clades. The avian genome is principally characterized by its constrained size, which predominantly arose because of lineage-specific erosion of repetitive elements, large segmental deletions, and gene loss. Avian genomes furthermore show a remarkably high degree of evolutionary stasis at the levels of nucleotide sequence, gene synteny, and chromosomal structure. Despite this pattern of conservation, we detected many non-neutral evolutionary changes in protein-coding genes and noncoding regions. These analyses reveal that pan-avian genomic diversity covaries with adaptations to different lifestyles and convergent evolution of traits.

With ~10,500 living species (*1*), birds are the most species-rich class of tetrapod vertebrates. Birds originated from a theropod lineage more than 150 million years ago during the Jurassic and are the only extant descendants of dinosaurs (*2*, *3*). The earliest diversification of extant birds (Neornithes) occurred during the Cretaceous period. However, the Neoaves, the most diverse avian clade, later underwent a rapid global expansion and radiation after a mass extinction event ~66 million years ago near the Cretaceous-Paleogene (K-Pg) boundary (*4*, *5*). As a result, the extant avian lineages exhibit extremely diverse morphologies and rates of diversification. Given the nearly complete global inventory of avian species, and the immense collected amount of distributional and biological data, birds are widely used as models for investigating evolutionary and ecological questions (*6*, *7*). The chicken (*Gallus gallus*), zebra finch (*Taeniopygia guttata*), and pigeon (rock dove) (*Columba livia*) are also important model organisms in disciplines such as neuroscience and developmental biology (*8*). In addition, birds are widely used for global conservation priorities (*9*) and are culturally important to human societies. A number of avian species have been domesticated and are economically important. Farmed and wild water birds are key players in the global spread of pathogens, such as avian influenza virus (*10*).

Despite the need to better understand avian genomics, annotated avian genomic data was previously available for only a few species: the domestic chicken, domestic turkey (*Meleagris gallopavo*) and zebra finch (*11*–*13*), together with a few others only published recently (*14*–*16*). To build an understanding of the genetic complexity of birds and to investigate links between their genomic variation and phenotypic diversity, we collected and compared genome sequences of these and other avian species (48 species total), representing all 32 neognath and two of the five palaeognath orders (Fig. 1) (*17*), thus representing nearly all of the major clades of living birds (*5*).

Results

Sequencing, assembly, and annotation

We used a whole-genome shotgun strategy to generate genome sequences of 45 new avian species (*18*), including two species representing two orders within the infraclass Paleognathae [common ostrich (*Struthio camelus*) and white-throated tinamou (*Tinamus guttatus*)], the other order within Galloanserae [Peking duck (*Anas platyrhynchos*)], and 41 species representing 30 neoavian orders (table S1) (*19*). In combination with the three previously published avian genomes (*11*–*13*), the genome assemblies cover 92% (34 of 37) of all avian orders (the three missing orders belong to the Paleognathae) (*17*). With the exception of the budgerigar (*Melopsittacus undulatus*), which was assembled through a multiplatform (Illumina/GS-FLX/PacBio) approach (*20*), all other new genomes were sequenced and assembled with Illumina (San Diego, CA) short reads (Fig. 1) (*18*). For 20 species, we produced high (>50×) coverage sequences from multiple libraries, with a gradient of insert sizes and built full-genome assemblies. For the remaining 25 species, we generated low (~30×) coverage data from two insert-size libraries and built less complete but still sufficient assemblies for comparative genome analyses. These de novo (*18*) genome assemblies ranged from 1.05 to 1.26 Gb, which is consistent with estimated cytology-based genome sizes (*21*), suggesting near complete genome coverage for all species. Scaffold N50 sizes for high-coverage genomes ranged from 1.2 to 6.9 Mb, whereas those for lower-coverage genomes were ~48 kb on average (table S2). The genomes of the ostrich and budgerigar were further assembled with optical maps, increasing their scaffold N50 sizes to 17.7 and 13.8 Mb, respectively (*20*, *22*).

We annotated the protein-coding sequences using a homology-based method for all genomes, aided by transcriptome sequencing for some species (*18*). To avoid systematic biases related to the use of different methods in annotations of previously published avian genomes, we created a uniform reference gene set that included all genes from the chicken, zebra finch, and human (*23*). This database was used to predict protein gene models in all avian genomes and American alligator (*Alligator mississippiensis*) (*24*). All high-coverage genomes were predicted to contain ~15,000 to 16,000 transposable element-free protein-coding genes [table S3 and annotation files in (*19*)], similar to the chicken genome (~15,000). Despite the fragmented nature of the low-coverage genomes leading to ~3000 genes likely missing or partially annotated, it was still possible to predict 70 to 80% of the entire catalog of avian genes.

Downloaded from www.sciencemag.org on December 14, 2014

Broad patterns of avian genome evolution

Although many fishes and some amphibians have smaller genomes than birds, among amniotes, birds have the smallest (*21*). The genomes of mammals and nonavian reptiles typically range from 1.0 to 8.2 Gb, whereas avian genomes range from 0.91 in the black-chinned hummingbird (*Archilochus alexanderi*) to a little over 1.3 Gb in the common ostrich (*21*). A number of hypotheses have been proposed for the smaller avian genome size (*25*–*28*). Here, we document key events that have likely contributed to this smaller genome size.

The proliferation and loss of transposable elements (TEs) may drive vertebrate genome size evolution (*29*–*31*). Consistent with the zebra finch and galliformes genomes (*11*–*13*, *32*), almost all avian genomes contained lower levels of repeat elements (~4 to 10% of each genome) (table S4) than in other tetrapod vertebrates (for example, 34 to 52% in mammals) (*33*). The sole outlier was the downy woodpecker (*Picoides pubescens*), with TEs representing ~22% of the genome, derived mainly from species-specific expansion of LINE (long interspersed elements) type CR1 (chicken repeat 1) transposons (fig. S1). In contrast, the average total length of SINEs (short interspersed elements) in birds has been reduced to ~1.3 Mb, which is ~10 to 27 times less than in other reptiles [12.6 Mb in alligator; 34.9 Mb in green sea turtle (*Chelonia mydas*)], suggesting that a deficiency of SINEs occurred in the common ancestor of birds.

We compared the average size of genomic elements of birds with 24 mammalian and the three nonavian reptile genomes. Avian protein-coding genes were on average 50 and 27% shorter than the mammalian and reptilian genes, respectively (Fig. 2A). This reduction is largely due to the shortening of introns and reduced intergenic distances that resulted in an increased gene density (Fig. 2A). Such genomic contraction has also evolved convergently in bats (fig. S11), the only flying mammalian group. The condensed genomes may represent an adaptation tied to rapid gene regulation required during powered flight (*34*, *35*).

To further investigate whether avian genome size reduction is due to a lineage-specific reduction in the common avian ancestor of birds or expansion in other vertebrates (*36*), we performed ancestral state reconstructions of small [<100 base pairs (bp)] deletion events across an alignment of four representative well-assembled avian and three reptile genomes (*18*) and found that the avian ancestral lineage experienced the largest number of small deletion events—about twice the number in the common ancestor of birds and crocodiles (fig. S12). In contrast, many fewer small deletion events occurred in modern avian lineages (fig. S12).

We next created a gene synteny map between the highest-quality assembled avian genome (ostrich) and other reptile genomes to document lineage-specific events of large segmental deletions (*18*). We detected 118 syntenic blocks, spanning a total of 58 Mb, that are present in alligator and turtle genomes but lost in all birds (table S8). In contrast, ~8x and ~5x fewer syntentic blocks were missing in alligator (14 blocks, 9 Mb) and turtle (27 blocks, 8 Mb) relative to green anole, respectively, confirming the polarity of genome size reduction in birds (table S8). The large segmental losses in birds were skewed to losses from chr2 and chr6 of the green anole (fig. S13). Two of the green anole's 12 pairs of microchromosomes, LGd and LGf, were completely missing in birds, with no homologous genes found within the avian genomes. Most of these lost segments were located at the ends of chromosomes or close to the centrosomes (fig. S13). Furthermore, lost segments were enriched at apparent breakpoints of the avian microchromosomes (Fig. 2B and fig. S13). These findings imply that the large segmental losses may be a consequence of chromosomal fragmentation events in the common ancestor of birds giving rise to additional microchromosomes in modern birds.

The large segmental deletions in birds contain at least 1241 functional protein-coding genes (table S9), with each lost segment containing at least five contiguous genes. The largest region lost in birds was a 2.1-Mb segment of the green anole chr2, which contains 28 protein-coding genes (Fig. 2B). Overall, at least 7% of the green anole macrochromosomal genes were lost through segmental deletions in birds. Although gene loss is a common evolutionary process, this massive level of segmental deletion has not been previously observed in vertebrates. Over 77% of the 1241 genes present in the large segmentally deleted regions have at least one additional paralog in the green anole genome, a level higher than the overall percentage of genes with paralogs in the green anole genome or avian genomes (both at ~70%). This suggests that birds may

[1]China National GeneBank, Beijing Genomics Institute (BGI)–Shenzhen, Shenzhen, 518083, China. [2]Centre for Social Evolution, Department of Biology, Universitetsparken 15, University of Copenhagen, DK-2100 Copenhagen, Denmark. [3]Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, 1350 Copenhagen, Denmark. [4]Royal Veterinary College, University of London, London, UK. [5]Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul 151-742, Republic of Korea. [6]Cho and Kim Genomics, Seoul National University Research Park, Seoul 151-919, Republic of Korea. [7]School of Biological Sciences, University of Nebraska, Lincoln, NE 68588, USA. [8]Centro de Investigación en Ciencias del Mar y Limnología (CIMAR)/Centro Interdisciplinar de Investigação Marinha e Ambiental (CIIMAR), Universidade do Porto, Rua dos Bragas, 177, 4050-123 Porto, Portugal. [9]Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, 4169-007 Porto, Portugal. [10]Department of Biological Sciences, University of South Carolina, Columbia, SC, USA. [11]Department of Biology and Molecular Biology, Montclair State University, Montclair, NJ 07043, USA. [12]Department of Animal Ecology, Uppsala University, Norbyvägen 18D, S-752 36 Uppsala, Sweden. [13]Marie Bashir Institute for Infectious Diseases and Biosecurity, Charles Perkins Centre, School of Biological Sciences and Sydney Medical School, The University of Sydney, Sydney, NSW 2006, Australia. [14]Program in Emerging Infectious Diseases, Duke-NUS Graduate Medical School, Singapore 169857, Singapore. [15]Department of Integrative Biology University of California, Berkeley, CA 94720, USA. [16]College of Life Sciences, Wuhan University, Wuhan 430072, China. [17]School of Bioscience and Bioengineering, South China University of Technology, Guangzhou 510006, China. [18]BGI Education Center,University of Chinese Academy of Sciences, Shenzhen, 518083, China. [19]Key Laboratory of Animal Models and Human Disease Mechanisms of Chinese Academy of Sciences and Yunnan Province, Kunming Institute of Zoology, Kunming, Yunnan 650223, China. [20]Department of Neurobiology, Howard Hughes Medical Institute, Duke University Medical Center, Durham, NC 27710, USA. [21]Institute of Biological, Environmental and Rural Sciences, Aberystwyth University, Aberystwyth, UK. [22]School of Biosciences, University of Kent, Canterbury CT2 7NJ, UK. [23]Instituto de Ciências Biomédicas Abel Salazar (ICBAS), Universidade do Porto, Portugal. [24]Department of Biology, University of California Riverside, Riverside, CA 92521, USA. [25]Department of Biochemistry, Molecular Biology, Entomology and Plant Pathology, Mississippi State University, Mississippi State, MS 39762, USA. [26]Institute for Genomics, Biocomputing and Biotechnology, Mississippi State University, Mississippi State, MS 39762, USA. [27]Instituto de Ciencias Ambientales y Evolutivas, Facultad de Ciencias, Universidad Austral de Chile, Valdivia, Chile. [28]Department of Anatomy, Physiology and Biochemistry, Swedish University of Agricultural Sciences, Post Office Box 7011, S-750 07, Uppsala, Sweden. [29]Department of Agricultural Biotechnology and Research Institute for Agriculture and Life Sciences, Seoul National University, Seoul 151-742, Republic of Korea. [30]State Key Laboratory for Agrobiotechnology, China Agricultural University, Beijing 100094, China. [31]College of Animal Science and Technology, China Agricultural University, Beijing 100094, China. [32]Organisms and Environment Division, Cardiff School of Biosciences, Cardiff University, Cardiff CF10 3AX, Wales, UK. [33]Key Lab of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101 China. [34]International Wildlife Consultants, Carmarthen SA33 5YL, Wales, UK. [35]Centre for Zoo and Wild Animal Health, Copenhagen Zoo, Roskildevej 38, DK-2000 Frederiksberg, Denmark. [36]Department of Ecology and Evolutionary Biology, Tulane University, New Orleans, LA, USA. [37]Museum of Natural Science, Louisiana State University, Baton Rouge, LA 70803, USA. [38]The Genome Institute at Washington University, St. Louis, MO 63108, USA. [39]College of Medicine and Forensics, Xi'an Jiaotong University, Xi'an, 710061, China. [40]Department of Biomolecular Engineering, University of California, Santa Cruz, CA 95064, USA. [41]Theodosius Dobzhansky Center for Genome Bioinformatics, St. Petersburg State University, St. Petersburg, Russia. [42]Nova Southeastern University Oceanographic Center 8000 N Ocean Drive, Dania, FL 33004, USA. [43]Smithsonian Conservation Biology Institute, National Zoological Park, 1500 Remount Road, Front Royal, VA 22630, USA. [44]Genetics Division, San Diego Zoo Institute for Conservation Research, 15600 San Pasqual Valley Road, Escondido, CA 92027, USA. [45]Department of Vertebrate Zoology, MRC-116, National Museum of Natural History, Smithsonian Institution, Post Office Box 37012, Washington, DC 20013-7012, USA. [46]Center for Macroecology, Evolution and Climate, the Natural History Museum of Denmark, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen O, Denmark. [47]Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, 1 Beichen West Road, Chaoyang District, Beijing 100101, China. [48]Swedish Species Information Centre, Swedish University of Agricultural Sciences, Box 7007, SE-750 07 Uppsala, Sweden. [49]Department of Biochemistry & Biophysics, University of California, San Francisco, CA 94158, USA. [50]Department of Organismic and Evolutionary Biology and Museum of Comparative Zoology, Harvard University, 26 Oxford Street, Cambridge, MA 02138, USA. [51]Department of Biology and Genetics Institute, University of Florida, Gainesville, FL 32611, USA. [52]Imperial College London, Grand Challenges in Ecosystems and the Environment Initiative, Silwood Park Campus, Ascot, Berkshire SL5 7PY, UK. [53]Division of Genetics and Genomics, The Roslin Institute and Royal (Dick) School of Veterinary Studies, The Roslin Institute Building, University of Edinburgh, Easter Bush Campus, Midlothian EH25 9RG, UK. [54]Department of Biology, New Mexico State University, Box 30001 MSC 3AF, Las Cruces, NM 88003, USA. [55]Macau University of Science and Technology, Avenida Wai long, Taipa, Macau 999078, China. [56]Trace and Environmental DNA Laboratory, Department of Environment and Agriculture, Curtin University, Perth, Western Australia, 6102, Australia. [57]Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, 2200 Copenhagen, Denmark. [58]Princess Al Jawhara Center of Excellence in the Research of Hereditary Disorders, King Abdulaziz University, Jeddah 21589, Saudi Arabia. [59]Department of Medicine, University of Hong Kong, Hong Kong.
*These authors contributed equally to this work. **†Corresponding author. E-mail: zhanggj@genomics.cn (G.Z.); jarvis@neuro.duke.edu (E.D.J.); mtpgilbert@gmail.com (M.T.P.G.); wangj@genomics.cn (J.W.)**
‡Present address: Department of Biological Sciences, Texas Tech University, Lubbock, TX 79409, USA. §Avian Genome Consortium authors and affiliations are listed at the end of this paper.

have undergone functional compensation in their paralogous gene copies, reducing selection against the loss of these segmental regions. We predict that the loss of functions associated with many genes in the avian ancestor may have had a profound influence on avian-specific traits (table S11).

Conservative mode of genome evolution

With ~2/3 of avian species possessing ~30 pairs of microchromosomes, the avian karyotype appears to be distinctly conserved because this phenotype is not a general feature of any other vertebrate group studied to date (*37*). We assessed the rates of avian chromosomal evolution among the 21 more fully assembled genomes (scaffold N50 > 1 Mb) (table S2) (*18*). From the alignment of chicken with the other 20 avian genomes, plus green anole and *Boa constrictor* (*38*), we identified homologous synteny blocks (HSBs) and 1746 evolutionary breakpoint regions (EBRs) in different avian lineages and then estimated the expected number of EBRs (*18*) and the rates of genomic rearrangements, using a phylogenetic total evidence nucleotide tree (TENT) as a guide (*5*). We excluded the turkey genome after detecting an unusually high fraction of small lineage-specific rearrangements, suggesting a high number of local misassemblies. Of the 18 remaining non–Sanger-sequenced genomes (table S2), the estimated rate of chimeric scaffolds that could lead to false EBRs was ~6% (*39*).

The average rate of rearrangements in birds is ~1.25 EBRs per million years; however, bursts of genomic reorganization occurred in several avian lineages (fig. S15). For example, the origin of Neognathae was accompanied by an elevated rate of chromosome rearrangements (~2.87 EBRs per million years). Intriguingly, all vocal learning species [zebra finch, medium-ground finch (*Geospiza fortis*), American crow (*Corvus brachyrhynchos*), budgerigar, and Anna's hummingbird (*Calypte anna*)] had significantly higher rates of rearrangements than those of close vocal nonlearning relatives [golden-collared manakin (*Manacus vitellinus*), peregrine falcon (*Falco peregrinus*) and chimney swift (*Chaetura pelagica*)] [phylogenetic analysis of variance F statistic (F) = 5.78, P = 0.0499] and even higher relative to all vocal nonlearning species (F = 15.03, P = 0.004). This may be related to the larger radiations these clades experienced relative to most other bird groups. However, the golden-collared manakin, which belongs to suboscines (vocal nonlearners) that have undergone a larger radiation than parrots and hummingbirds, has a low rearrangement rate.

We next compared microsynteny (local gene arrangements), which is more robust and accurate than macrosynteny analyses for draft assemblies (*18*). We compared with eutherian mammals, which are approximately the same evolutionary age as Neoaves and whose genome assemblies are of similar quality. We examined the fraction of orthologous genes identified from each pair of two-avian/mammalian genomes, on the basis of syntenic and best reciprocal blast matches (*18*). Birds have a significantly higher percentage of synteny-defined orthologous genes than that of mammals (Fig. 2C). The fraction of genes retained in syntenic blocks in any pairwise comparison was linearly related with evolutionary time, by which the overall level of genome shuffling in birds was lower than in mammals over the past ~100 million years (Fig. 2C). This suggests a higher level of constraint on maintaining gene synteny in birds relative to mammals.

The apparent stasis in avian chromosome evolution suggests that birds may have experienced relatively low rates of gene gain and loss in multigene families. We examined the intensively studied gene families that encode the various α- and β-type subunits of hemoglobin, the tetrameric protein responsible for blood oxygen transport in jawed vertebrates (*40*). In amniotes, the α- and β-globin gene families are located on different chromosomes (*40*) and experienced high rates of gene turnover because of lineage-specific duplication and deletion events (*41*). In birds, the size and membership composition of the globin gene families have remained remarkably constant during ~100 million years of evolution, with most examined species retaining an identical complement (Fig. 2D). Estimated gene turnover rates (λ) of α- and β-globin gene families were over twofold higher in mammals than birds (λ = 0.0023 versus 0.0011, respectively). Much of the variation in the avian α-globin gene family was attributable to multiple independent inactivations of the α*D*-globin gene (Fig. 2D), which encodes the α-chain subunit of a

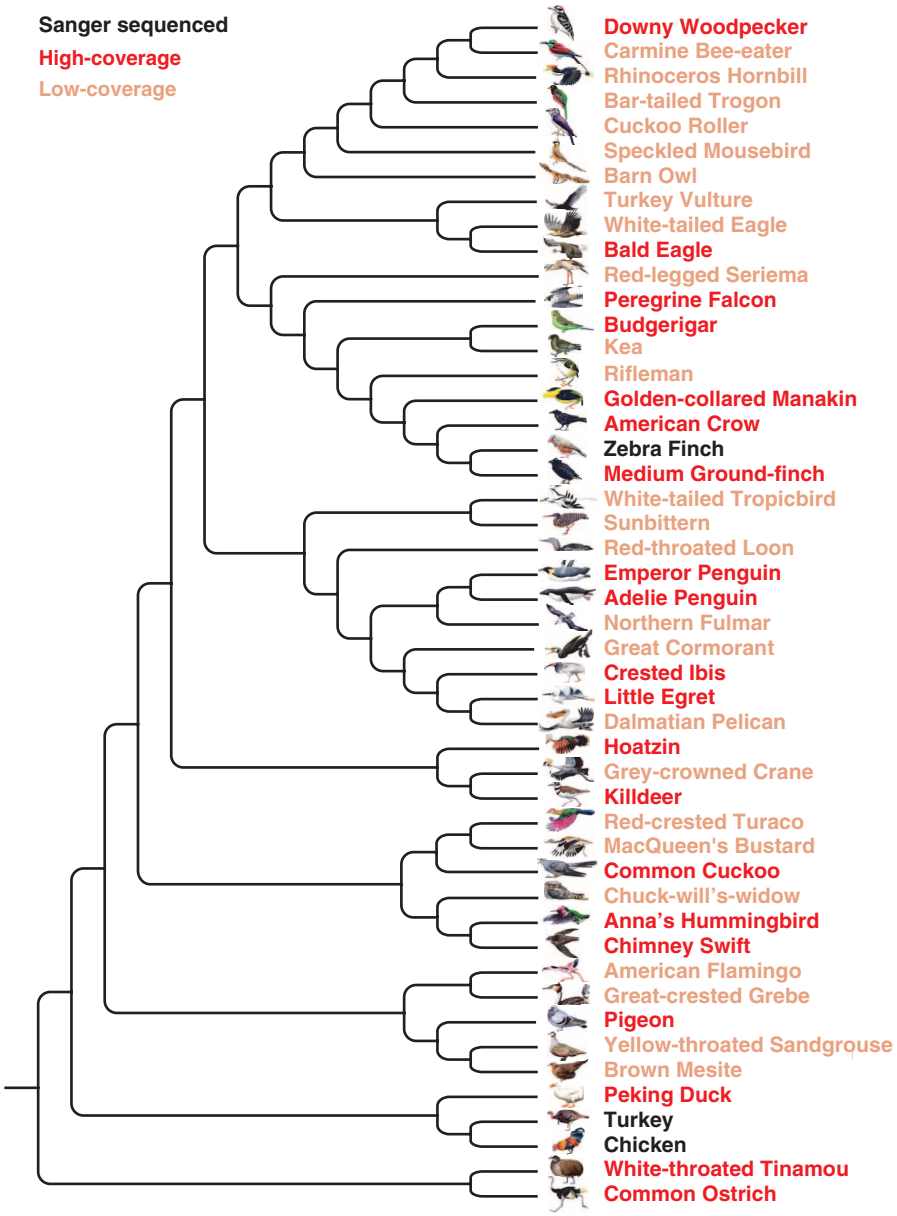


Fig. 1. Avian family tree and genomes sequenced. The phylogenomic relationships of the 48 avian genomes from (*5*), with Sanger-sequenced (black), high-coverage (dark red), and low-coverage (light red) genomes denoted.

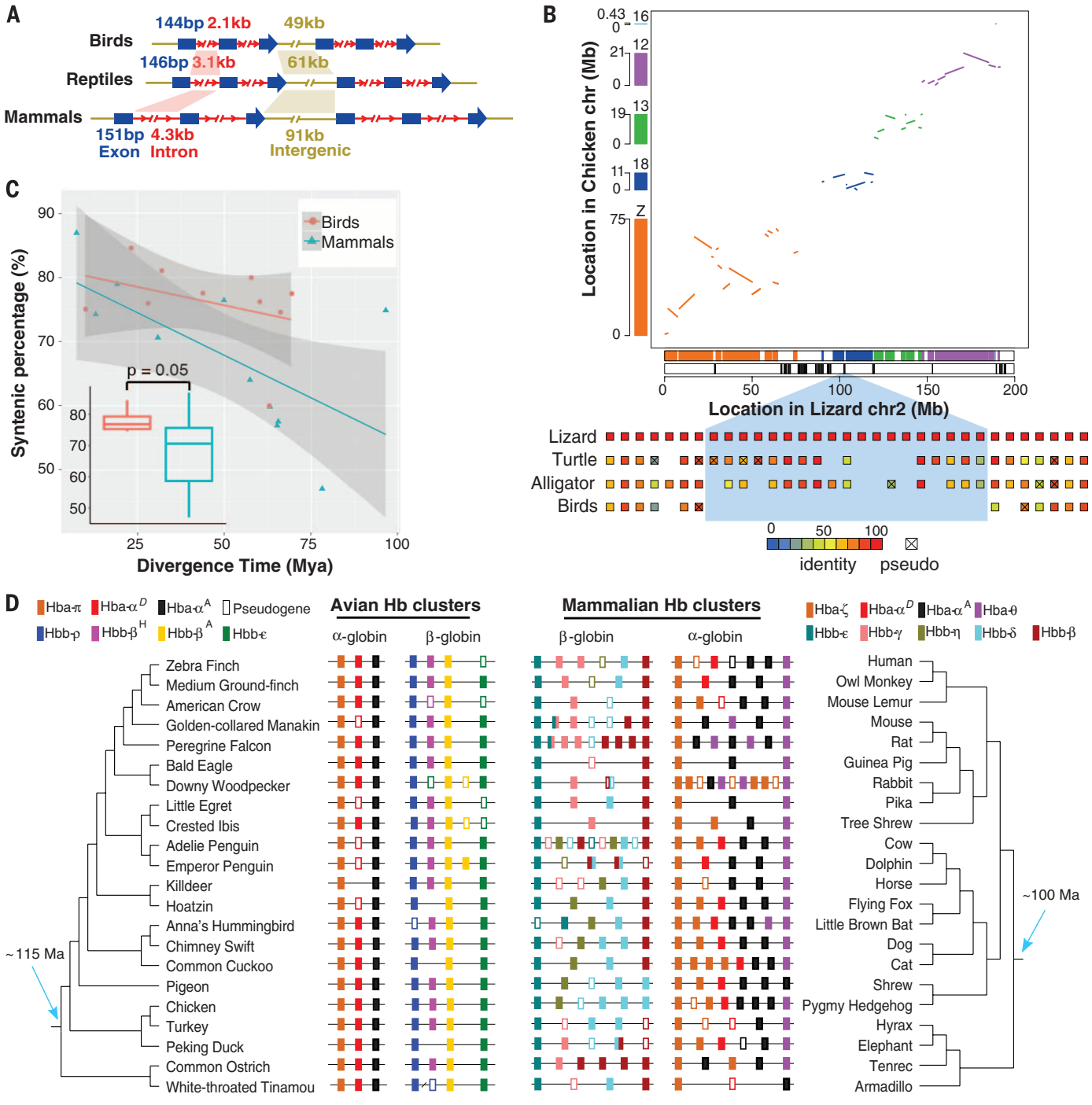


Fig. 2. Genome reduction and conservation in birds. (**A**) Comparison of average size of introns, exons, and intergenic regions within avian, reptilian, and mammalian genomes. (**B**) Synteny plot and large segmental deletions between green anole chromosome 2 and multiple chicken chromosomes. Colored bars and lines indicate homologous blocks between two species; black bars indicate location of large avian-specific segmental deletions, which are enriched at the breakpoints of interchromosome rearrangements. (Bottom) An example of a large segmental deletion in birds (represented by ostrich genes). Homologous genes annotated in each species are shown in small boxes. The color spectrum represents the percent identity of homologous genes with the green anole. (**C**) Distribution of gene synteny percentages identified for phylogenetically independent species pairs of various divergence ages. Dots indicate the percentage of genes remaining in a syntenic block in pairwise comparisons between two avian or mammalian species. Box plots indicate that the overall distributions of the synteny percentages in birds and mammals are different (*P* value was calculated by using Wilcoxon rank sum test with phylogenetically independent species pairs). (**D**) Chromosomal organization of the α- and β-globin gene clusters in representative avian and mammalian taxa. These genes encode the α- and β-type subunits of tetrameric (α2β2) hemoglobin isoforms that are expressed at different ontogenetic stages. In the case of the α-like globin genes, birds and mammals share orthologous copies of the *αD*- and *αA*-globin genes. Likewise, the avian π-globin and the mammalian ζ-globin genes are 1:1 orthologs. In contrast, the genes in the avian and mammalian β-globin gene clusters are derived from independent duplications of one or more β-like globin genes that were inherited from the common ancestor of tetrapod vertebrates (*90*, *91*).

hemoglobin isoform (HbD) expressed in both embryonic and definitive erythrocytes (*42*). Because of uniform and consistent differences in oxygen-binding properties between HbD and the major adult-expressed hemoglobin isoform, HbA (which incorporates products of αA-globin) (*42*), the inactivations of αD-globin likely contribute to variation in blood-oxygen affinity, which has important consequences for circulatory oxygen transport and aerobic energy metabolism. Overall, the globin gene families illustrate a general pattern of evolutionary stasis in birds relative to mammals.

Genomic nucleotide substitution rates vary across species and are determined through both neutral and adaptive evolutionary processes (*43*). We found that the overall pan-genomic background substitution rate in birds (~1.9×10^{-3} substitutions per site per million years) was lower than in mammals (~2.7×10^{-3} substitution per site per million years) (Fig. 3A). However, the substitution rate estimates also exhibited interordinal variation among birds (Fig. 3A). There was a positive correlation between the substitution rate and the number of species per order [coefficient of determination (R^2) = 0.21, P = 0.01, Pearson's test with phylogenetically independent contrasts] (Fig. 3B and fig. S19), evidencing an association with rates of macroevolution (*44*). For example, Passeriformes, the most diverse avian order, exhibited the highest evolutionary rate (~3.3×10^{-3} substitutions per site per million years), almost two times the average of Neoaves (~2×10^{-3} substitutions per site per million years, Fig. 3A). Landbirds exhibited an average higher substitution rate than that of waterbirds (landbirds, ~2.2×10^{-3} substitutions per site per million years; waterbirds, ~1.6×10^{-3} substitutions per site per million years), which is consistent with the observation that landbirds have greater net diversification rates than those of waterbirds (*7*). Among the landbirds, the predatory lineages exhibited slower rates of evolution (~1.6×10^{-3} substitutions per site per million years), similar to that of waterbirds. Moreover, the three vocal learning landbird lineages (parrots, songbirds, and hummingbirds) are evolving faster than are nonvocal learners (Fig. 3A). Overall, our analyses indicate that genome-wide variation in rates of substitution is a consequence of the avian radiation into a wide range of niches and associated phenotypic changes.

Selective constraints on functional elements

Conservation of DNA sequences across distantly related species reflects functional constraints (*45*). A direct comparison of 100-Mb orthologous genomic regions revealed more regions evolving slower than the neutral rate among birds (Fig. 3C) than mammals (*46*), which is consistent with the slower rate of avian mitochondrial sequence evolution (*47*). We predicted 3.2 million highly conserved elements (HCEs) at a resolution of 10 bp or greater spanning on average 7.5% of the avian genome, suggesting a strong functional constraint in avian genomes. Functional annotations revealed that ~12.6% of these HCEs were associated with protein-coding genes, whereas

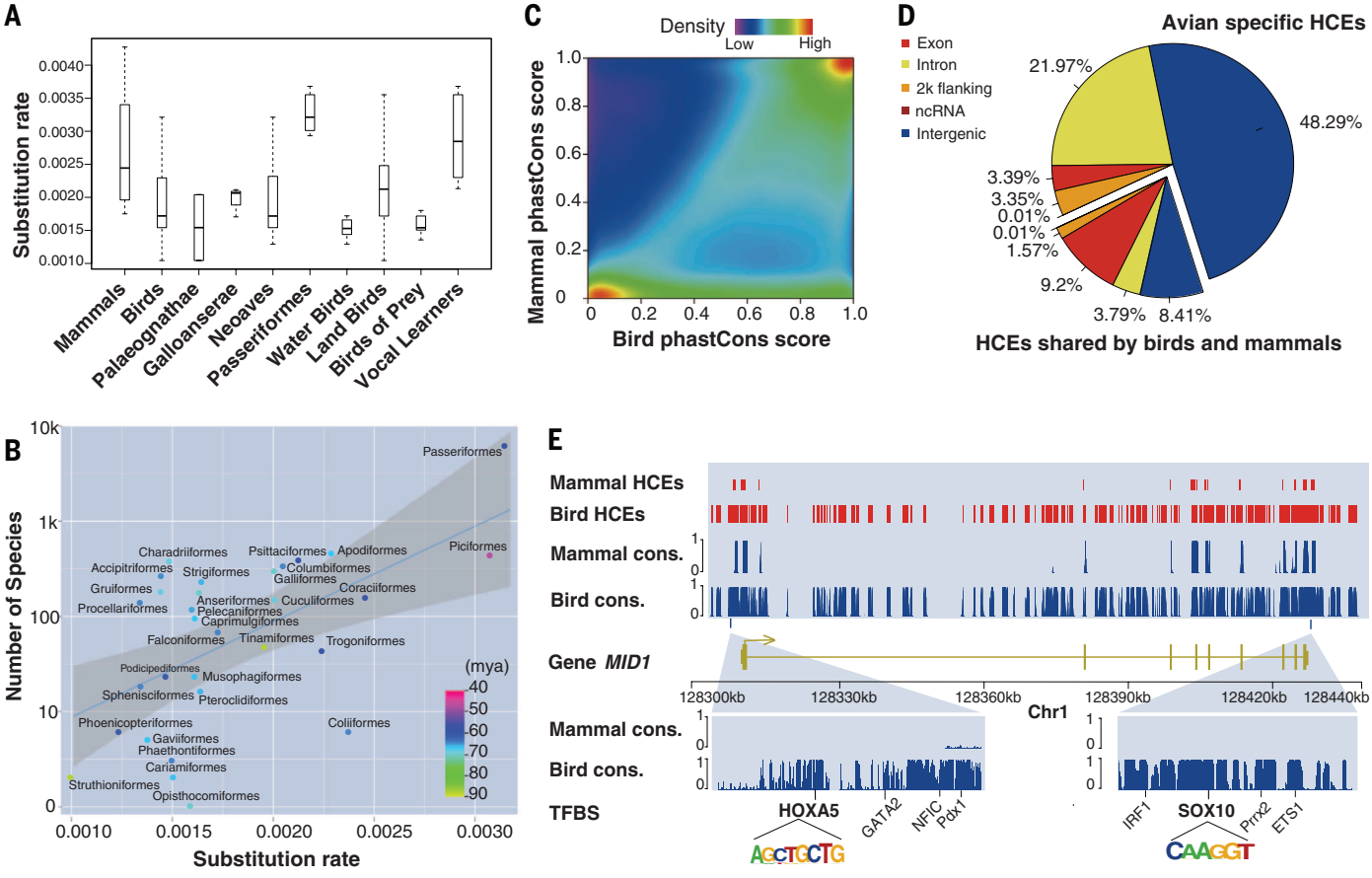


Fig. 3. Evolutionary rate and selection constraints. (**A**) Substitution rate in each lineage was estimated by the comparison of fourfold degenerate (4d) sites in coding regions, in units of substitutions per site per million years. Waterbirds and landbirds are defined in (*5*). (**B**) Correlation between average substitution rates and number of species within different avian orders. Divergence times were estimates from (*5*). The fit line was derived from least square regression analysis, and the confidence interval was estimated by "stat_smooth" in R. The units of the *x* axis are numbers of substitutions per site per million years. The correlation figure with phylogenetically independent contrasts is provided in the supplementary materials. (**C**) Density map for comparison of conservation levels between pan-avian and pan-mammalian genomes, on the basis of the homologous genomic regions between birds and mammals. Conservation levels were quantified by means of PhastCons basewise conservation scores. (**D**) HCEs found in both mammalian and avian genomes (smaller pie piece) and those that are avian-specific (larger pie piece). (**E**) MID1 contains abundant avian-specific HCEs in the upstream and downstream regulatory regions. Many regulatory motif elements are identified in these avian-specific HCEs. Cons., conservation level.

the majority of the remaining HCEs were located in intron and intergenic regions (Fig. 3, D and E). These HCEs enabled us to identify 717 new protein-coding exons and 137 new protein-coding genes, with 77% of the latter supported by the deep transcriptome data (table S17). Deep transcriptome sequencing also enabled us to annotate 5879 candidate long noncoding RNA (lncRNA) genes, of which 220 overlapped HCEs with a coverage ratio of >50% (table S18) (*18*).

Because HCEs may have different functions in different lineages, we separated the HCEs into two categories: bird-specific and amniote HCEs (shared by birds and mammals). Among the bird-specific HCEs, we identified 13 protein-coding genes that were highly conserved in birds but divergent in mammals (table S19). One of the most conserved was the sperm adhesion gene, *SPAM1*, which mediates sperm binding to the egg coat (*48*). This gene, however, was under positive selection driven by sperm competition in mammalian species (*49*). Noncoding HCEs play important roles in the regulation of gene expression (*50*); thus, we compared the transcription factor binding sites in the ENCODE project (*51*) with the HCEs and found that the avian-specific HCEs are significantly associated with transcription factors functioning in metabolism (table S20), whereas amniote core HCEs are enriched with transcription factors functioning in signal regulation, stimulus responses, and development (table S21).

To investigate evolutionary constraints on gene regions, we calculated d*N*/d*S* [the ratio of the number of nonsynonymous substitutions per nonsynonymous site (dN) to the number of synonymous substitutions per synonymous site (d*S*)] for 8295 high-quality orthologs. Consistent with the fast-Z sex chromosome hypothesis (*52*), the evolutionary rate of Z-linked genes was significantly higher than autosome genes (Fig. 4A). This is most likely driven by the reduction of effective population size (N_e) of Z-linked genes—because the N_e of Z chromosome is only 3/4 of that of autosomes—as well as by male sexual selection (*52*). Furthermore, consistent with the fast-macro hypothesis, the overall rate of macrochromosomal genic evolution is higher than that of microchromosomes (Fig. 4A), which is probably due to differences in the recombination rates and genic densities between macro- and microchromosomes in birds (*53*).

We also examined the d*N*/d*S* ratio of each avian Gene Ontology (GO) category for comparison with mammals and within birds. Those involved in development (such as spinal cord development and bone resorption) are evolving faster in birds, and those involved in the brain function (such as synapse assembly, synaptic vesicle transport, and neural crest cell migration) are evolving faster in mammals (tables S23 and S24). Genes involved in oxidoreductase activity were relatively rapidly evolving in the Palaeognathae clade that contains the flightless ratites (Fig. 4B and table S25). The fast evolving GOs in the Galloanserae participate in regulatory functions (Fig. 4B and table S26). In Neoaves, genes involved in microtubule-based processes were the fastest evolving (Fig. 4B and table S27). We speculate that these differences could be caused by relaxed selective constraints or positive selection in different lineages.

Genotype-phenotype convergent associations: Evolution of vocal learning

With the availability of genomes representing all major modern avian lineages and their revised phylogenetic relationships (*5*), it becomes possible to conduct genome-wide association studies across species with convergent traits. We focused on vocal learning, which given our phylogenetic analyses is inferred as having evolved independently, either twice, in hummingbirds and the common ancestor of songbirds and parrots, or three times (*5*, *54*). All three groups have specialized song-learning forebrain circuits (song nuclei) not found in vocal nonlearners (Fig. 5A) (*55*).

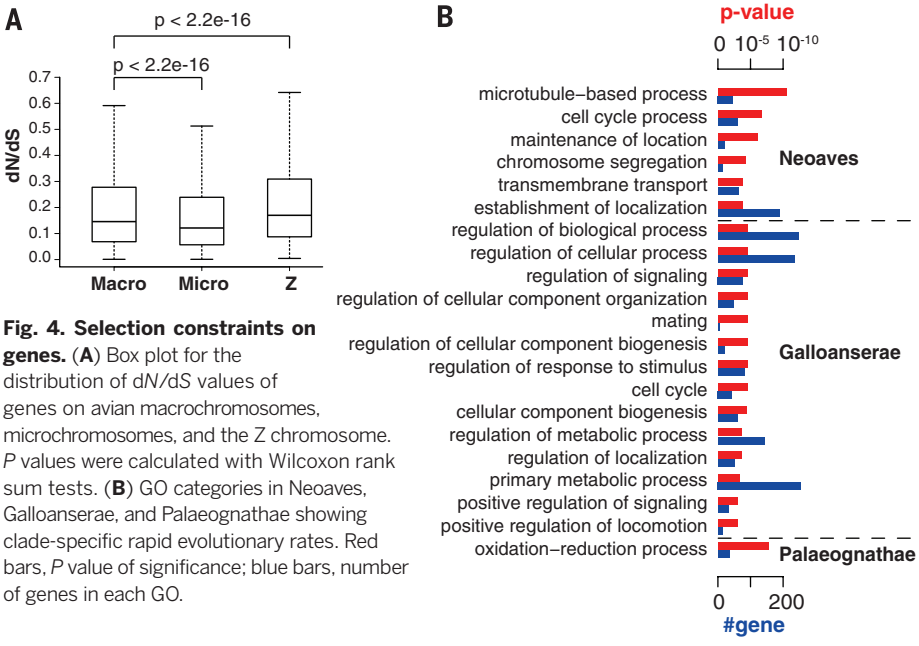


Fig. 4. Selection constraints on genes. (**A**) Box plot for the distribution of d*N*/d*S* values of genes on avian macrochromosomes, microchromosomes, and the Z chromosome. *P* values were calculated with Wilcoxon rank sum tests. (**B**) GO categories in Neoaves, Galloanserae, and Palaeognathae showing clade-specific rapid evolutionary rates. Red bars, *P* value of significance; blue bars, number of genes in each GO.

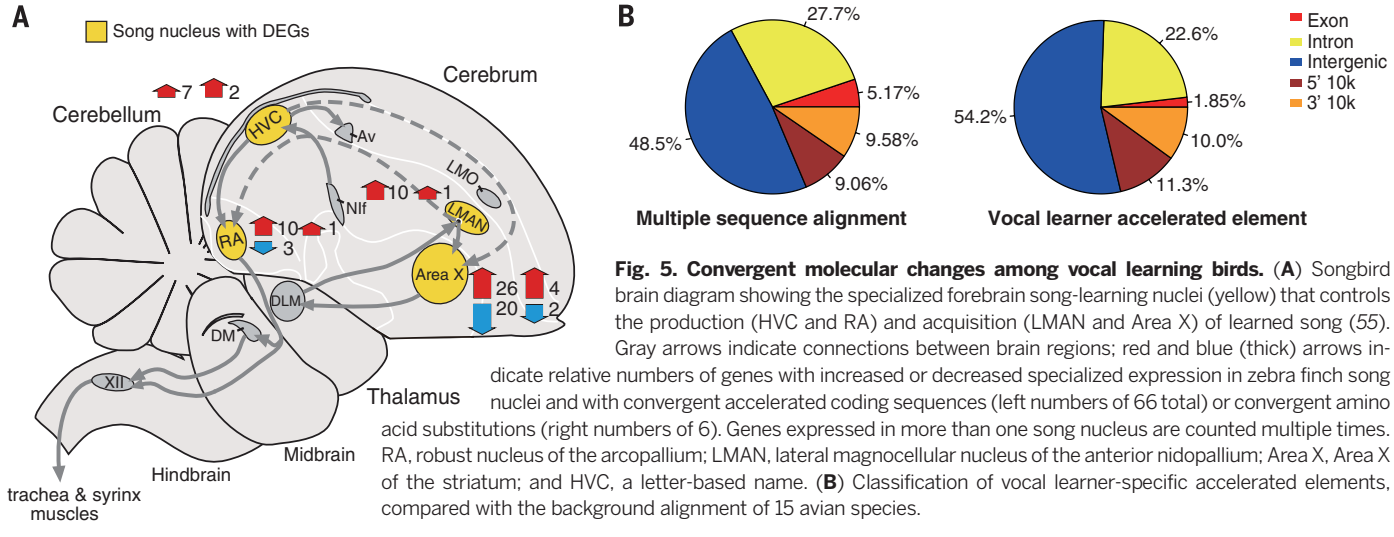


Fig. 5. Convergent molecular changes among vocal learning birds. (**A**) Songbird brain diagram showing the specialized forebrain song-learning nuclei (yellow) that controls the production (HVC and RA) and acquisition (LMAN and Area X) of learned song (*55*). Gray arrows indicate connections between brain regions; red and blue (thick) arrows indicate relative numbers of genes with increased or decreased specialized expression in zebra finch song nuclei and with convergent accelerated coding sequences (left numbers of 66 total) or convergent amino acid substitutions (right numbers of 6). Genes expressed in more than one song nucleus are counted multiple times. RA, robust nucleus of the arcopallium; LMAN, lateral magnocellular nucleus of the anterior nidopallium; Area X, Area X of the striatum; and HVC, a letter-based name. (**B**) Classification of vocal learner-specific accelerated elements, compared with the background alignment of 15 avian species.

Analyses of 7909 orthologous protein-coding genes with available amino acid sites in all three vocal-learning and control vocal nonlearning groups revealed convergent accelerated d*N*/d*S* for 227 genes in vocal learners (table S28). Of these, 73% (165) were expressed in the songbird brain (physically cloned mRNAs), and of these, 92% (151) were expressed in adult song-learning nuclei, which is much higher than the expected 60% of brain genes expressed in song nuclei (*56*). About 20% (*33*) were regulated by singing, which is twice the expected 10% (*56*). In addition, 41% of the song nuclei accelerated genes showed differential expression among song nuclei [expected 20% (*56*)], and 0.7 to 9% [0.7 to 4.3% expected (*57*)] showed specialized expression compared with the surrounding brain regions (table S28) (*58*). GO analyses of the accelerated differentially expressed song nuclei genes revealed 30 significant functionally enriched gene sets, which clustered into four major categories, including neural connectivity, brain development, and neural metabolism (fig. S25). For an independent measure of convergence, we developed an approach that scans for single amino acid substitutions common to species with a shared trait, controlling for phylogenetic relationships (*18*). Of the 7909 genes, 38 had one to two amino acid substitutions present only in vocal learners (table S31). At least 66% of these were expressed in the songbird brain, including in the song nuclei [58%; 20% expected (*56*)]. Two genes (*GDPD4* and *KIAA1919*) showed convergent accelerated evolution on the amino acid sites specific to vocal learners (table S31).

To identify accelerated evolution in noncoding sequences in vocal learners, we scanned the genome alignment using phyloP (*18*, *59*). We used a more limited sampling of vocal nonlearning species closely related to the vocal learners (table S32) because of the relatively faster evolutionary rate of noncoding regions. We scanned the entire genome alignment and found 822 accelerated genomic elements specifically shared by all three vocal learning groups (table S33). These convergent elements were skewed to intergenic regions in vocal learners relative to the background average accelerated elements across species (Fisher's exact test, $P < 2.2 \times 10^{-16}$) (Fig. 5B). Of these elements, 332 were associated with 278 genes (within 10 kb 5′ or 3′ of the nearest gene), of which a high proportion (76%) was expressed in the brain; almost all of those (94%, 198 genes) expressed in one or more song nuclei, 20% were regulated by singing (10% expected), 51% (20% expected) showed differential expression among song nuclei, and 2 to 15% [0.7 to 4.3% expected, based on (*56*)] had specialized expression relative to the surrounding brain regions, including the *FoxP1* gene involved in speech (table S34 and figs. S27 to S32). Overall, these analyses show a 2- to 3.5-fold enrichment of accelerated evolution in regulatory regions of genes differentially expressed in vocal learning brain regions. In contrast, there was very little overlap (2.5%) of genes with convergent accelerated noncoding changes and convergent accelerated amino acid changes, indicating two independent targets of selection for convergent evolution.

Evolution of ecologically relevant genes

We also investigated candidate genes that underlie traits relevant to avian ecological diversity. Although these analyses should be approached with caution given the phenotypic and ecological plasticity within major avian lineages, we examined genes putatively associated with major skeletal and tissue changes for the capacity for powered flight, feeding modification such as loss of teeth, the advanced visual system found in some lineages, and sexual and reproductive systems.

Evolution of the capacity for flight

Skeletal systems: The evolution of flight involved a series of adaptive changes at the morphological and molecular levels. One of the key requirements for flight is a skeleton that is both strong and lightweight. In both birds and nonavian theropods, this evolved through the fusion and elimination of some bones and the pneumatization of the remaining ones (*60*). Of 89 genes involved in ossification (table S36), 49 (~55%) showed evidence of positive selection in birds, which is almost twice as high as in mammals (31 genes, ~35%). For birds, most of these are involved in the regulation of remodeling and ossification-associated processes, or bone development in general, and those with the highest values for global d*N*/d*S* (>0.5) were obtained for *AHSG* (*α-2-HS-glycoprotein*), which is associated with bone mineral density, and *P2RX7* (*P2X purinoceptor 7*), which is associated with bone homeostasis. The variation in the extension of pneumatization in avian post-cranial bones has been associated with the variation in body size and foraging strategies (*61*). Therefore, selection of these genes may explain variation in the levels of bone pneumatization in birds because the genes involved in the process of maintaining trabeculae within bones likely depends on the intrinsic network of genes participating in bone resorption and mineralization. These results suggest that most structural differences in bone between birds and mammals may be a result of bone remodeling and resorption (table S37).

Pulmonary structure and function: The increased metabolism associated with homeothermy and powered flight requires an efficient gas exchange process during pulmonary ventilation. Because of functional integration of ventilation and locomotion, birds evolved a volume-constant lung and a rigid trunk region, whereas mammals evolved a changing-volume lung, often coupled to locomotory flexion of the lumbar region (*62*). In contrast to the pulmonary alveola of the mammalian lung, the avian lung has a honeycomb-like structure incorporating a flow-through system with small air capillaries (*63*). We found five genes that function in mammalian lung development that were lost in the avian ancestor (table S11).

Feathers: The evolution and subsequent morphological diversification of feathers have shaped avian physiology, locomotion, mate choice, and ecological niches (*64*). Feathers are composed of α- and β-keratins (*65*), the latter of which are structural proteins found only in the epidermal appendages of birds and other reptiles. The α-keratin gene family has contracted in birds relative to reptiles (except turtle) and mammals (0.7-fold change), whereas the β-keratin gene family has expanded (1.96-fold change) relative to reptiles (Fig. 6A and table S39). The avian β-keratins form six clusters, with all major avian lineages possessing members from each avian cluster (fig. S33), indicating that avian β-keratin diversity was present in the basal avian lineage. Of these, the feather β-keratin subfamily is avian-specific and comprises over 56% of the genes, whereas the remaining avian β-keratin subfamilies (claw, scale, and keratinocyte β-keratin subfamilies) are found in turtles and crocodiles (Fig. 6A and fig. S33). The mean number of keratinocyte β-keratins is similar across bird groups and their two closest living reptile relatives (turtle and alligator), suggesting copy number conservation since their common ancestor (Fig. 6A). In contrast, aquatic/semi-aquatic birds have a relatively low mean number of feather β-keratins compared with that of land birds, with land birds having more than double the number, and among them several domesticated land birds (zebra finch, chicken, pigeon, and budgerigar) having more than 8 times (Fig. 6A). Although the later observation is concordant with the hypothesis that domestication may increase the recombination rate at β-keratin loci (*66*, *67*), domestic turkey and Peking duck did not exhibit this trend. Overall, these findings indicate that feather compositional adaptations are associated with different avian lifestyles.

Evolution of genes related to diet

Edentulism: The evolution of birds also had major consequences with regard to their feeding strategies and diets, with changes at the structural, biochemical, and sensory levels (among others). One of the most immediately obvious avian-specific traits is edentulism, the phenotype of being toothless. Edentulism is thought to have evolved independently in multiple theropod lineages (*68*). However, although most phylogenetic analyses suggest that teeth were lost in the common ancestor of modern birds (*69*), several studies have recovered dentate taxa (*Hesperornis* and *Ichthyornis*) from the Mesozoic inside of crown Neornithes, suggesting that tooth loss could have occurred independently (*70*). A scan of avian genomes for molecular fossils of tooth-specific genes recovered remnants of enamel and dentin formation genes in all species examined [table 1 in (*71*)]. Frameshift mutations and whole-exon deletions were widespread in all investigated tooth genes. The vast majority of debilitating mutations were not shared, but all species shared unambiguous deletions in protein-coding exons of enamel-specific genes (*ENAM*, *AMEL*, *AMBN*, *MMP20*, and *AMTN*) and one dentin-specific gene (*DSPP*). This shared pattern of pseudogenization across living birds supports the hypothesis that the common ancestor of modern birds lacked mineralized teeth (*69*).

Diet-related enzymes: Birds have evolved an extraordinary diversity of dietary specializations. The glyoxylate detoxifying enzyme alanine/glyoxylate aminotransferase (AGT) represents a candidate for study (*72*). We recovered complete AGT genes from 22 avian genomes (table S42), of which five exhibit pseudogenized forms in their MTS region (Fig. 6B and fig. S34). MTS function was lost in three unrelated avian orders, which is consistent with multiple independent dietary transitions during avian evolution. Detection of positively selected amino acids at 137 Q ($dN/dS = 2.153$) and 378 R ($dN/dS = 2.153$) in all birds provided additional support for diet-related adaptation in AGT (positions according to human AGT; posterior probability > 99%; $P < 0.0001$).

Vitamin C (Vc) is an important nutrient cofactor in a range of essential metabolic reactions. Loss of the ability to synthesize Vc has occurred in humans, Guinea pigs, and some bats. All species that do not synthesize Vc exhibit a pseudogenized gene for L-gulonolactone oxidase (*GULO*), an enzyme essential for catalyzing the last step of Vc synthesis (*73*). Genomic mining revealed *GULO* pseudogenization in two oscines (medium ground-finch and zebra finch) and the suboscine golden-collared manakin (Fig. 6B and fig. S35). In contrast, intact *GULO* was recovered from the third oscine species, American crow, and the basal passerine rifleman (*Acanthisitta chloris*) (table S43). Similar to mammals (*74*), this pseudogenization was caused by the loss of different exons and lethal mutations. We also found purifying selection has dominated *GULO* evolution, from the ancestral amniote node ($dN/dS = 0.096$) to ancestral birds ($dN/dS = 0.133$) and mammals ($dN/dS = 0.355$), suggesting conservation of the ability to synthesize Vc both before and after avian divergence. However, both the American crow and rifleman exhibited nonsynonymous changes in *GULO* at one order of magnitude higher than the average (fig. S36), a sign of potentially harmful mutations (*75*).

Rhodopsin/opsins and vision

Birds exhibit what is possibly the most advanced vertebrate visual system, with a highly developed ability to distinguish colors over a wide range of wavelengths. In contrast to mammals, which have relatively few photoreceptor classes, almost all birds studied to date have retained an ancestral tetrapod set of cones hypothesized to play a role in reproduction and feeding (*76*). Vertebrate visual opsins are classified into five genes in two families: rhodopsin (RH1) and conopsins (RH2, OPN1sw1, OPN1sw2, and OPN1lw). In most avian genomes, we detected higher numbers of opsin genes than in mammalian genomes, which lacked *OPN4x* (*77*), *RH2*, and either *OPN1sw1* (Monotremata) or *OPN1sw2* (Theria). All avian genomes contained *RH1* and *RH2*, and most high-coverage genomes contained two to three of the remaining three conopsin genes (table S44), supporting that ancestral avian vision is tetrachromatic. Penguins were one of the exceptions, with both species exhibiting only three classes of functional opsins, and thus are trichromatic, which is in line with retinal examination (*78*). This is likely due to their aquatic lifestyle and is consistent with observations of marine mammals that also appear to have lost one, or even both, cone pigment (or pigments) (*76*).

Signs of strong positive selection were detected in the branch leading to the passerine group Passerida (represented by the medium ground-finch and zebra finch) (fig. S38), which corroborates that the shift from violet sensitive SWS1 cones in this clade was adaptive (*79*). Excluding these species, dN/dS values for *OPN1sw1* were lower in birds than in mammals (Fig. 6C). Optimal color discrimination requires an even distribution of spectral sensitivities (*80*), which is more easily disturbed with an increasing number of cone classes. Hence, stabilizing selection on spectral sensitivity should be stronger in birds than in mammals, and the dN/dS values are consistent with this prediction. Besides two transmembrane regions (II and VII) encompassing previously identified spectral tuning amino acids in the SWS1 conopsin, we found markedly positive selection in region IV, strongly suggesting that there is one or more unknown amino acid sites important to spectral tuning of this ultraviolet-sensitive cone (fig. S40).

Sex-related and reproductive traits

Reproduction-related genes: Unlike other reptiles, almost all birds develop only a single functional ovary, on the left side (*81*), as a result of the evolutionary loss of the right ovary during the transition from nonavian theropods to birds (*82*). It has been hypothesized that this loss represents an adaptation to reduce weight during flight (*82*). We found that two genes related with ovary development (*MMP19* and *AKR1C3*) have been lost in birds. *MMP19*, a matrix metalloprotease gene, functions during the follicular growth and the ovulation process (*83*), and the enzyme AKR1C3 catalyzes the conversion of androstenedione to testosterone and has been associated with polycystic ovary syndrome (*84*).

We analyzed a range of other genes related to reproduction, under the hypothesis that some of them may have been direct targets of the morphological and behavioral adaptations related to sexual selection in birds. Reproduction genes in *Drosophila*, humans, and marine invertebrates evolve faster than do nonreproduction genes

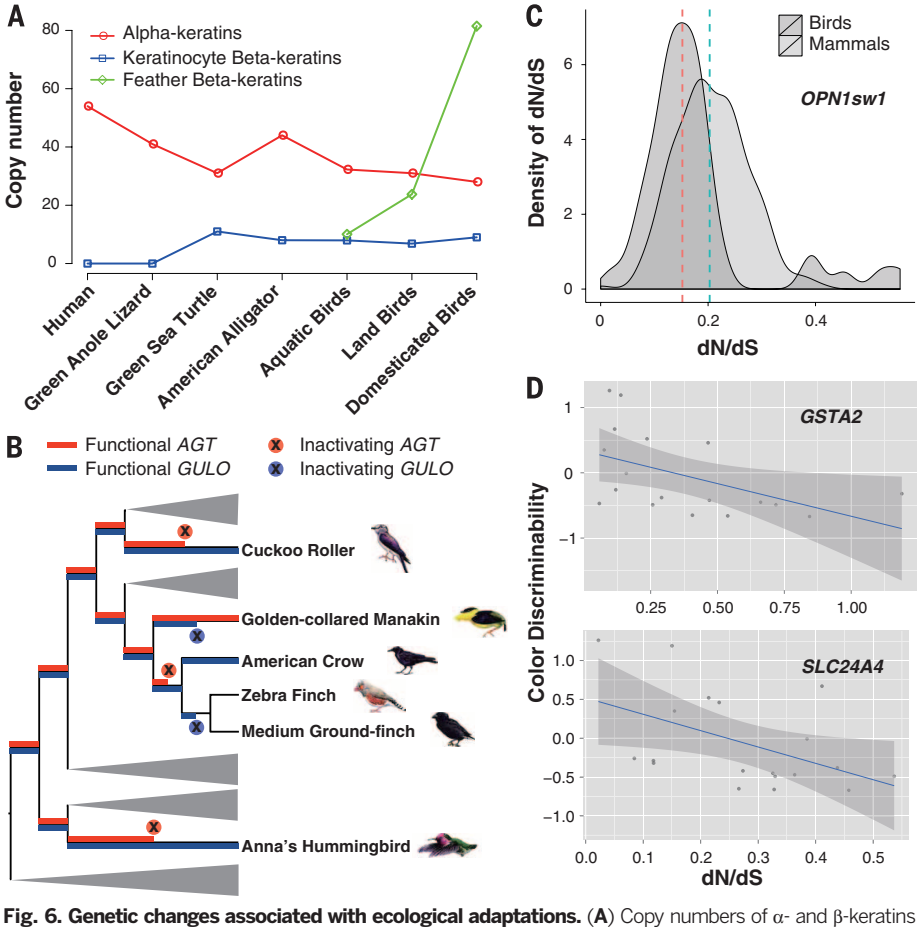


Fig. 6. Genetic changes associated with ecological adaptations. (**A**) Copy numbers of α- and β-keratins in humans, reptiles, and birds, including in aquatic birds, land birds, and domesticated birds. Definitions of aquatic and landbirds are provided in (*5*). (**B**) Pseudogenization events of the diet-related genes *AGT* and *GULO* along the avian phylogeny. (**C**) Density distribution of dN/dS values of the *OPN1sw1* gene for mammals (median, 0.21) and birds (median, 0.16). (**D**) dN/dS values of two plumage color–related genes (*GSTA2* and *SLC24A4*) show negative correlation with the color discriminability values (log transformation applied). The correlation figures with phylogenetically independent contrasts are provided in the supplementary materials.

(*49*). We chose 89 genes that may be involved in spermatogenesis (table S46) and six involved in oogenesis (table S47). We found that 19 out of 46 avian species show significantly accelerated evolution (lineage-specific d*N*/d*S* ratio) of spermatogenesis genes relative to the genomic background (table S48). In contrast, only the carmine bee-eater (*Merops nubicoides*) and Peking duck showed significantly accelerated evolution in oogenesis genes (table S49). These results suggest that male birds are the dominant targets of sexual selection, which drives rapid evolution of spermatogenesis genes via sperm competition (*85*).

Plumage color: We investigated the genomics of plumage color, a behaviorally important trait and longstanding example of sexual selection (*86*). Male birds have frequently evolved extravagant plumage color in response to both male-male competition and female choice (*87*, *88*), resulting in remarkable sexual dichromatism. Analysis of 15 genes implicated in avian plumage coloration demonstrated rapid evolutionary rates over the genomic average in 8 of 46 lineages (table S51). This pattern suggests that these genes are evolving under adaptive evolution.

Carotenoids, which are responsible for the bright yellow and red pigments that underlie some of the most conspicuous coloration patterns in vertebrates, unlike melanins can be only acquired through diet and represent trade-offs between coloration and other physiological conditions. We identified a negative correlation between color discriminability and d*N*/d*S* across birds for the gene *GSTA2*, which is involved in the binding and deposition of carotenoids and in plumage dichromatism ($R^2 = 0.24$, $P = 0.045$, Pearson's test with phylogenetically independent contrasts) (Fig. 6D and fig. S41), and similarly for *SLC24A4*, which is associated with hair color in humans ($R^2 = 0.21$, $P = 0.056$, Pearson's test with phylogenetically independent contrasts) (Fig. 6D and fig. S42), suggesting that either diversifying and stabilizing selection or the effect of different population sizes is driving the evolution of plumage color genes.

Discussion and conclusions

The small genome size of birds with fragmented microchromosomes and reduced repeat transposon activity, in contrast to other vertebrates, has been a static feature in the avian clade for >100 million years. Avian genomes consistently contain fewer genes, ~70% of the number of the human genome, and with one detected exception (downy woodpecker), an extremely reduced fraction of repeat elements. Thus, the ancestral avian lineage has distinctly lost a large number of genes by means of large segmental deletions after their divergence from other extant reptiles. These large genomic sequence deletions appear to be linked to a second defining feature of avian genomes: the putatively ancestral fission of macrochromosomes into a relatively large number of microchromosomes.

Genome conservation in birds—along with regard to sequence, synteny, and chromosomal structure—is remarkable in light of their rapid historical radiation. This is considerably different from the evolution of mammalian genomes, which although are experiencing a rapid radiation at a similar time, today display richer genome shuffling and variation (*89*). By comparing the genomes of 48 birds that are constrained within a largely resolved phylogeny, we discovered millions of highly constrained elements comprising 7.5% of avian genomes. This evolutionary profiling of genomes across >100 million years (*5*) enables their interpretation in a functional genomic context not possible in previous genomic studies restricted to fewer taxa.

The analyses of genome sequences for taxa distributed across the avian phylogeny also explains the rich biodiversity of the avian clade because we identified selective constraints on certain categories of genes in different avian lineages. Convergent evolution also appears to be shaping the evolution of protein-coding genes and their regulatory elements, establishing similar morphological or behavioral features in distantly related bird species, as well as variation in specific gene families that correspond to avian traits and environmental adaptation. We believe that the data and analyses presented here open a new window into the evolution, diversification, and ecological adaptation of tetrapod vertebrates and offers a phylogenomic perspective that helps bridge the chasm between micro- and macroevolution.

REFERENCES AND NOTES

1. F. Gill, D. Donsker, IOC World Bird List (version 3.5) (2013).
2. L. M. Chiappe, L. M. Witmer, *Mesozoic Birds: Above the Heads of Dinosaurs* (Univ. California Press, Berkeley, CA, 2002).
3. G. Dyke, G. W. Kaiser, *Living Dinosaurs the Evolutionary History of Modern Birds* (Wiley-Blackwell, Hoboken, NJ, 2011).
4. A. Feduccia, *Trends Ecol. Evol.* **18**, 172–176 (2003).
5. E. D. Jarvis *et al*., *Science* **346**, 1320–1331 (2014).
6. B. G. Holt *et al*., *Science* **339**, 74–78 (2013).
7. W. Jetz, G. H. Thomas, J. B. Joy, K. Hartmann, A. O. Mooers, *Nature* **491**, 444–448 (2012).
8. H. Zeigler, P. E. Marler, in Behavioral Neurobiology of Birdsong, December 2002, Hunter College, City University of New York, New York, NY, US; this volume is the result of the aforementioned conference which was one of an annual symposium series sponsored by the Hunter College Gene Center. (2004).
9. A. J. Stattersfield, M. J. Crosby, A. J. Long, D. C. Wege, *Endemic Bird Areas of the World: Priorities for Conservation. BirdLife Conservation* (BirdLife International, Cambridge, UK, 1998).
10. D. J. Alexander, *Vet. Microbiol.* **74**, 3–13 (2000).
11. L. W. Hillier *et al*., *Nature* **432**, 695–716 (2004).
12. R. A. Dalloul *et al*., *PLOS Biol.* **8**, e1000475 (2010).
13. W. C. Warren *et al*., *Nature* **464**, 757–762 (2010).
14. Y. Huang *et al*., *Nat. Genet.* **45**, 776–783 (2013).
15. X. Zhan *et al*., *Nat. Genet.* **45**, 563–566 (2013).
16. M. D. Shapiro *et al*., *Science* **339**, 1063–1067 (2013).
17. E. Dickinson, J. Remsen, *The Howard and Moore Complete Checklist of the Birds of the World* (Aves Press, Eastbourne, UK, 2013).
18. Materials and methods are available as supplementary materials on *Science* Online.
19. G. Zhang *et al*., *GigaScience* **3**, 26 (2014).
20. G. Ganapathy *et al*., *GigaScience* **3,** 11 (2013).
21. T. R. Gregory, in The Animal Genome Size Database (2005); available at www.genomesize.com.
22. Q. Zhou *et al*., *Science* **346**, 1246338 (2014).
23. E. S. Lander *et al*., *Nature* **409**, 860–921 (2001).
24. R. E. Green *et al*., *Science* **346**, 1254449 (2014).
25. A. L. Hughes, M. K. Hughes, *Nature* **377**, 391 (1995).
26. M. Rho *et al*., *Genome Biol. Evol.* **1**, 2–12 (2009).
27. S. Morand, R. E. Ricklefs, *Trends Genet.* **17**, 567–568 (2001).
28. E. Waltari, S. V. Edwards, *Am. Nat.* **160**, 539–552 (2002).
29. M. G. Kidwell, *Genetica* **115**, 49–63 (2002).
30. C. Feschotte, E. J. Pritham, *Annu. Rev. Genet.* **41**, 331–368 (2007).
31. M. Lynch, J. S. Conery, *J. Struct. Funct. Genomics* **3**, 35–44 (2003).
32. N. Sela, E. Kim, G. Ast, *Genome Biol.* **11**, R59 (2010).
33. A. Böhne, F. Brunet, D. Galiana-Arnoux, C. Schultheis, J. N. Volff, *Chromosome Res.* **16**, 203–215 (2008).
34. Q. Zhang, S. V. Edwards, *Genome Biol. Evol.* **4**, 1033–1043 (2012).
35. C. L. Organ, A. M. Shedlock, A. Meade, M. Pagel, S. V. Edwards, *Nature* **446**, 180–184 (2007).
36. A. L. Hughes, R. Friedman, *Mol. Biol. Evol.* **25**, 2681–2688 (2008).
37. D. K. Griffin, L. B. W. Robertson, H. G. Tempest, B. M. Skinner, *Cytogenet. Genome Res.* **117**, 64–77 (2007).
38. K. R. Bradnam *et al*., *GigaScience* **2**, 10 (2013).
39. J. Kim *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **110**, 1785–1790 (2013).
40. F. G. Hoffmann, J. C. Opazo, J. F. Storz, *Mol. Biol. Evol.* **29**, 303–312 (2012).
41. F. G. Hoffmann, J. C. Opazo, J. F. Storz, *Mol. Biol. Evol.* **25**, 591–602 (2008).
42. M. T. Grispo *et al*., *J. Biol. Chem.* **287**, 37647–37658 (2012).
43. C. F. Baer, M. M. Miyamoto, D. R. Denver, *Nat. Rev. Genet.* **8**, 619–631 (2007).
44. C. Venditti, M. Pagel, *Trends Ecol. Evol.* **25**, 14–20 (2010).
45. A. Siepel *et al*., *Genome Res.* **15**, 1034–1050 (2005).
46. K. Lindblad-Toh *et al*., *Nature* **478**, 476–482 (2011).
47. B. Nabholz, S. Glémin, N. Galtier, *BMC Evol. Biol.* **9**, 54 (2009).
48. W. F. Lathrop, E. P. Carmichael, D. G. Myles, P. Primakoff, *J. Cell Biol.* **111**, 2939–2949 (1990).
49. W. J. Swanson, V. D. Vacquier, *Nat. Rev. Genet.* **3**, 137–144 (2002).
50. L. A. Pennacchio *et al*., *Nature* **444**, 499–502 (2006).
51. ENCODE Project Consortium, *Nature* **489**, 57–74 (2012).
52. J. E. Mank, E. Axelsson, H. Ellegren, *Genome Res.* **17**, 618–624 (2007).
53. E. Axelsson, M. T. Webster, N. G. C. Smith, D. W. Burt, H. Ellegren, *Genome Res.* **15**, 120–125 (2005).
54. A. Suh *et al*., *Nat. Commun.* **2**, 443 (2011).
55. E. D. Jarvis, *Ann. N. Y. Acad. Sci.* **1016**, 749–777 (2004).
56. O. Whitney *et al*., *Science* **346**, 1256780 (2014).
57. A. R. Pfenning *et al*., *Science* **346**, 1256846 (2014).
58. H. J. Karten *et al*., *J. Comp. Neurol.* **521**, 3702–3715 (2013).
59. M. J. Hubisz, K. S. Pollard, A. Siepel, *Brief. Bioinform.* **12**, 41–51 (2011).
60. E. R. Dumont, *Proc. Biol. Sci.* **277**, 2193–2198 (2010).
61. S. C. Gutzwiller, A. Su, P. M. O'Connor, *Anat. Rec.* **296**, 867–876 (2013).
62. H. R. Duncker, *Respir. Physiol. Neurobiol.* **144**, 111–124 (2004).
63. J. B. West, R. R. Watson, Z. Fu, *Respir. Physiol. Neurobiol.* **157**, 382–390 (2007).
64. F. Gill, *Ornithology* (W.H. Freeman and Company, New York, 1995).
65. A. R. Haake, G. König, R. H. Sawyer, *Dev. Biol.* **106**, 406–413 (1984).
66. J. Ross-Ibarra, *Am. Nat.* **163**, 105–112 (2004).
67. A. Blirt, G. Bell, *Nature* **326**, 803–805 (1987).
68. A. Louchart, L. Viriot, *Trends Ecol. Evol.* **26**, 663–673 (2011).
69. B. C. Livezey, R. L. Zusi, *Zool. J. Linn. Soc.* **149**, 1–95 (2007).
70. J. Cracraft, *Syst. Biol.* **31**, 35–56 (1982).
71. R. W. Meredith, G. Zhang, M. T. P. Gilbert, E. D. Jarvis, M. S. Springer, *Science* **346**, 1254390 (2014).
72. G. M. Birdsey, J. Lewin, A. A. Cunningham, M. W. Bruford, C. J. Danpure, *Mol. Biol. Evol.* **21**, 632–646 (2004).
73. I. B. Chatterjee, *Science* **182**, 1271–1272 (1973).
74. J. Cui, X. Yuan, L. Wang, G. Jones, S. Zhang, *PLOS ONE* **6**, e27114 (2011).
75. J. Cui, Y.-H. Pan, Y. Zhang, G. Jones, S. Zhang, *Mol. Biol. Evol.* **28**, 1025–1031 (2011).
76. W. I. L. Davies, S. P. Collin, D. M. Hunt, *Mol. Ecol.* **21**, 3121–3158 (2012).
77. S. S. Pires *et al*., *Proc. Biol. Sci.* **274**, 2791–2799 (2007).
78. J. K. Bowmaker, G. R. Martin, *J. Comp. Physiol. A Neuroethol. Sens. Neural Behav. Physiol.* **156**, 71–77 (1985).
79. A. Ödeen, O. Håstad, P. Alström, *BMC Evol. Biol.* **11**, 313 (2011).
80. L. Chittka, *Naturwissenschaften* **83**, 136–138 (1996).
81. S. S. Guraya, *Ovarian Follicles in Reptiles and Birds* (Springer-Verlag, Berlin, Germany, 1989).
82. X. Zheng *et al*., *Nature* **495**, 507–511 (2013).
83. M. Jo, T. E. Curry Jr., *Biol. Reprod.* **71**, 1796–1806 (2004).
84. M. O. Goodarzi, H. J. Antoine, R. Azziz, *J. Clin. Endocrinol. Metab.* **92**, 2659–2664 (2007).
85. T. R. Birkhead, A. P. Møller, Sperm competition and sexual selection. (Academic Press, 1998).
86. C. Darwin, *The Descent of Man, and Selection in Relation to Sex* (J. Murray, London, 1871).
87. M. Zuk, J. D. Ligon, R. Thornhill, *Anim. Behav.* **44**, 999–1006 (1992).
88. C. Mateos, J. Carranza, *Anim. Behav.* **54**, 1205–1214 (1997).
89. A. R. Quinlan, I. M. Hall, *Trends Genet.* **28**, 43–53 (2012).

90. J. F. Storz, J. C. Opazo, F. G. Hoffmann, *Mol. Phylogenet. Evol.* **66**, 469–478 (2013).
91. F. G. Hoffmann, J. F. Storz, T. A. Gorr, J. C. Opazo, *Mol. Biol. Evol.* **27**, 1126–1138 (2010).

ACKNOWLEDGMENTS

Genome assemblies and annotations of avian genomes in this study are available on the avian phylogenomics website (http://phybirds.genomics.org.cn), GigaDB (http://dx.doi.org/10.5524/101000), National Center for Biotechnology Information (NCBI), and ENSEMBL (NCBI and Ensembl accession numbers are provided in table S2). The majority of this study was supported by an internal funding from BGI. In addition, G.Z. was supported by a Marie Curie International Incoming Fellowship grant (300837); M.T.P.G. was supported by a Danish National Research Foundation grant (DNRF94) and a Lundbeck Foundation grant (R52-A5062); C.L. and Q.L. were partially supported by a Danish Council for Independent Research Grant (10-081390); and E.D.J. was supported by the Howard Hughes Medical Institute and NIH Directors Pioneer Award DP1OD000448.

The Avian Genome Consortium

Chen Ye,[1] Shaoguang Liang,[1] Zengli Yan,[1] M. Lisandra Zepeda,[2] Paula F. Campos,[2] Amhed Missael Vargas Velazquez,[2] José Alfredo Samaniego,[2] María Avila-Arcos,[2] Michael D. Martin,[2] Ross Barnett,[2] Angela M. Ribeiro,[3] Claudio V. Mello,[4] Peter V. Lovell,[4] Daniela Almeida,[3,5] Emanuel Maldonado,[3] Joana Pereira,[3] Kartik Sunagar,[3,5] Siby Philip,[3,5] Maria Gloria Dominguez-Bello,[6] Michael Bunce,[7] David Lambert,[8] Robb T. Brumfield,[9] Frederick H. Sheldon,[9] Edward C. Holmes,[10] Paul P. Gardner,[11] Tammy E. Steeves,[11] Peter F. Stadler,[12] Sarah W. Burge,[13] Eric Lyons,[14] Jacqueline Smith,[15] Fiona McCarthy,[16] Frederique Pitel,[17] Douglas Rhoads,[18] David P. Froman[19]

[1]China National GeneBank, BGI-Shenzhen, Shenzhen 518083, China. [2]Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, 1350 Copenhagen, Denmark. [3]CIMAR/CIIMAR, Centro Interdisciplinar de Investigação Marinha e Ambiental, Universidade do Porto, Rua dos Bragas, 177, 4050-123 Porto, Portugal. [4]Department of Behavioral Neuroscience Oregon Health & Science University Portland, OR 97239, USA. [5]Departamento de Biologia, Faculdade de Ciências, Universidade do Porto, Rua do Campo Alegre, 4169-007 Porto, Portugal. [6]Department of Biology, University of Puerto Rico, Av Ponce de Leon, Rio Piedras Campus, JGD 224, San Juan, PR 009431-3360, USA. [7]Trace and Environmental DNA laboratory, Department of Environment and Agriculture, Curtin University, Perth, Western Australia 6102, Australia. [8]Environmental Futures Research Institute, Griffith University, Nathan, Queensland 4121, Australia. [9]Museum of Natural Science, Louisiana State University, Baton Rouge, LA 70803, USA. [10]Marie Bashir Institute for Infectious Diseases and Biosecurity, Charles Perkins Centre, School of Biological Sciences and Sydney Medical School, The University of Sydney, Sydney NSW 2006, Australia. [11]School of Biological Sciences, University of Canterbury, Christchurch 8140, New Zealand. [12]Bioinformatics Group, Department of Computer Science, and Interdisciplinary Center for Bioinformatics, University of Leipzig, Hrtelstrasse 16-18, D-04107 Leipzig, Germany. [13]European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, Cambridge CB10 1SD, UK. [14]School of Plant Sciences, BIO5 Institute, University of Arizona, Tucson, AZ 85721, USA. [15]Division of Genetics and Genomics, The Roslin Institute and Royal (Dick) School of Veterinary Studies, The Roslin Institute Building, University of Edinburgh, Easter Bush Campus, Midlothian EH25 9RG, UK. [16]Department of Veterinary Science and Microbiology, University of Arizona, 1117 E Lowell Street, Post Office Box 210090-0090, Tucson, AZ 85721, USA. [17]Laboratoire de Génétique Cellulaire, INRA Chemin de Borde-Rouge, Auzeville, BP 52627 , 31326 CASTANET-TOLOSAN CEDEX, France. [18]Department of Biological Sciences, Science and Engineering 601, University of Arkansas, Fayetteville, AR 72701, USA. [19]Department of Animal Sciences, Oregon State University, Corvallis, OR 97331, USA.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/346/6215/1311/suppl/DC1
Supplementary Text
Figs. S1 to S42
Tables S1 to S51
References (*92*–*192*)

27 January 2014; accepted 6 November 2014
10.1126/science.1251385

RESEARCH ARTICLE

Whole-genome analyses resolve early branches in the tree of life of modern birds

Erich D. Jarvis,[1]*† Siavash Mirarab,[2]* Andre J. Aberer,[3] Bo Li,[4,5,6] Peter Houde,[7] Cai Li,[4,6] Simon Y. W. Ho,[8] Brant C. Faircloth,[9,10] Benoit Nabholz,[11] Jason T. Howard,[1] Alexander Suh,[12] Claudia C. Weber,[12] Rute R. da Fonseca,[6] Jianwen Li,[4] Fang Zhang,[4] Hui Li,[4] Long Zhou,[4] Nitish Narula,[7,13] Liang Liu,[14] Ganesh Ganapathy,[1] Bastien Boussau,[15] Md. Shamsuzzoha Bayzid,[2] Volodymyr Zavidovych,[1] Sankar Subramanian,[16] Toni Gabaldón,[17,18,19] Salvador Capella-Gutiérrez,[17,18] Jaime Huerta-Cepas,[17,18] Bhanu Rekepalli,[20] Kasper Munch,[21] Mikkel Schierup,[21] Bent Lindow,[6] Wesley C. Warren,[22] David Ray,[23,24,25] Richard E. Green,[26] Michael W. Bruford,[27] Xiangjiang Zhan,[27,28] Andrew Dixon,[29] Shengbin Li,[30] Ning Li,[31] Yinhua Huang,[31] Elizabeth P. Derryberry,[32,33] Mads Frost Bertelsen,[34] Frederick H. Sheldon,[33] Robb T. Brumfield,[33] Claudio V. Mello,[35,36] Peter V. Lovell,[35] Morgan Wirthlin,[35] Maria Paula Cruz Schneider,[36,37] Francisco Prosdocimi,[36,38] José Alfredo Samaniego,[6] Amhed Missael Vargas Velazquez,[6] Alonzo Alfaro-Núñez,[6] Paula F. Campos,[6] Bent Petersen,[39] Thomas Sicheritz-Ponten,[39] An Pas,[40] Tom Bailey,[41] Paul Scofield,[42] Michael Bunce,[43] David M. Lambert,[16] Qi Zhou,[44] Polina Perelman,[45,46] Amy C. Driskell,[47] Beth Shapiro,[26] Zijun Xiong,[4] Yongli Zeng,[4] Shiping Liu,[4] Zhenyu Li,[4] Binghang Liu,[4] Kui Wu,[4] Jin Xiao,[4] Xiong Yinqi,[4] Qiuemei Zheng,[4] Yong Zhang,[4] Huanming Yang,[48] Jian Wang,[48] Linnea Smeds,[12] Frank E. Rheindt,[49] Michael Braun,[50] Jon Fjeldsa,[51] Ludovic Orlando,[6] F. Keith Barker,[52] Knud Andreas Jønsson,[51,53,54] Warren Johnson,[55] Klaus-Peter Koepfli,[56] Stephen O'Brien,[57,58] David Haussler,[59] Oliver A. Ryder,[60] Carsten Rahbek,[51,54] Eske Willerslev,[6] Gary R. Graves,[51,61] Travis C. Glenn,[62] John McCormack,[63] Dave Burt,[64] Hans Ellegren,[12] Per Alström,[65,66] Scott V. Edwards,[67] Alexandros Stamatakis,[3,68] David P. Mindell,[69] Joel Cracraft,[70] Edward L. Braun,[71] Tandy Warnow,[2,72]† Wang Jun,[48,73,74,75,76]† M. Thomas P. Gilbert,[6,43]† Guojie Zhang[4,77]†

To better determine the history of modern birds, we performed a genome-scale phylogenetic analysis of 48 species representing all orders of Neoaves using phylogenomic methods created to handle genome-scale data. We recovered a highly resolved tree that confirms previously controversial sister or close relationships. We identified the first divergence in Neoaves, two groups we named Passerea and Columbea, representing independent lineages of diverse and convergently evolved land and water bird species. Among Passerea, we infer the common ancestor of core landbirds to have been an apex predator and confirm independent gains of vocal learning. Among Columbea, we identify pigeons and flamingoes as belonging to sister clades. Even with whole genomes, some of the earliest branches in Neoaves proved challenging to resolve, which was best explained by massive protein-coding sequence convergence and high levels of incomplete lineage sorting that occurred during a rapid radiation after the Cretaceous-Paleogene mass extinction event about 66 million years ago.

The diversification of species is not always gradual but can occur in rapid radiations, especially after major environmental changes (*1*, *2*). Paleobiological (*3*–*7*) and molecular (*8*) evidence suggests that such "big bang" radiations occurred for neoavian birds (e.g., songbirds, parrots, pigeons, and others) and placental mammals, representing 95% of extant avian and mammalian species, after the Cretaceous to Paleogene (K-Pg) mass extinction event about 66 million years ago (Ma). However, other nuclear (*9*–*12*) and mitochondrial (*13*, *14*) DNA studies propose an earlier, more gradual diversification, beginning within the Cretaceous 80 to 125 Ma. This debate is confounded by findings that different data sets (*15*–*19*) and analytical methods (*20*, *21*) often yield contrasting species trees. Resolving such timing and phylogenetic relationships is important for comparative genomics, which can inform about human traits and diseases (*22*).

Recent avian studies based on fragments of 5 [~5000 base pairs (bp) (*8*)] and 19 [31,000 bp (*17*)] genes recovered some relationships inferred from morphological data (*15*, *23*) and DNA-DNA hybridization (*24*), postulated new relationships, and contradicted many others. Consistent with most previous molecular and contemporary morphological studies (*15*), they divided modern birds (Neornithes) into Palaeognathae (tinamous and flightless ratites), Galloanseres [Galliformes (landfowl) and Anseriformes (waterfowl)], and Neoaves (all other extant birds). Within Neoaves,

they proposed several large new clades, including a waterbird clade containing taxa such as penguins, pelicans, and loons, as well as a landbird clade containing woodpeckers, birds of prey, parrots, and songbirds. Despite these efforts, the relationships among the deepest branches within Neoaves; the positions of a number of chronically challenging taxa such as shorebirds, mousebirds, owls, and the enigmatic hoatzin; and the identification of the first divergence of Neoaves [proposed to have given rise to two equally large clades designated Metaves and Coronaves (*25*)] remain unresolved.

Although some of the findings of the initial multigene studies (*8*, *17*) have since been corroborated with larger sequence (*26*–*28*) or transposable element (TE) insertion data sets (*29*), other proposed clades were not supported (*27*, *28*). Furthermore, complete mitochondrial genome analyses recover different relationships (*14*, *18*) and fail to support higher landbird monophyly [but see (*30*)]. Some of the differences among studies could arise from gene tree incongruence, possibly due to incomplete lineage sorting (ILS) of those genes (*29*, *31*), nucleotide base composition biases (*19*), differences between data types (*32*, *33*), or insufficient data (*34*, *35*). Thus, it has been difficult to establish confidence in whether specific avian traits—such as vocal learning, predatory behavior, or adaptations to aquatic or terrestrial habitats—reflect single or multiple independent origins and under what ecological conditions these events have occurred.

A common assumption is that whole-genome data will improve phylogenetic reconstructions, due to the complete evolutionary record within each species' genome and increased statistical power (*34*, *35*). We test this hypothesis through phylogenetic analysis on 48 avian genomes we collected or assembled, representing all commonly accepted extant neognath orders (*36*, *37*) and two palaeognaths, with several nonavian reptiles and human as outgroups.

Species choice, computational developments, and total evidence nucleotide data set

We chose species representing all neoavian orders according to different classifications [see supplementary materials section 1 (SM1)]. These include groups that have been challenging to place within the avian tree, such as the hoatzin, cuckoo-roller, nightjars, mousebirds, mesites, and seriemas (table S1). We also included species postulated to descend from deep nodes in their orders to break up potentially long branches, such as the kea for parrots (Psittaciformes) and the rifleman for songbirds (Passeriformes). We included vocal-learning species (oscine songbirds, hummingbirds, and parrots), used as models for spoken language in humans (*38*), and their proposed closest vocal-nonlearning relatives (suboscines, swifts, falcons, and/or cuckoos, depending on the tree) to help resolve differences in trees that lead to different conclusions on their independent gains (*15*, *17*, *18*, *26*, *29*, *38*, *39*). The resulting data set consisted of 45 avian genomes sequenced in part for this project [48 when including previously published species (*40*–*42*)] and three nonavian reptiles [American alligator, green sea turtle, and green anole lizard (*43*)] (table S1), with details reported in (*44*–*52*).

We were confronted with computational challenges not previously encountered in smaller-scale phylogenomic studies. Differently annotated genomes complicated the identification of orthologs, and the size of the data matrix made it impossible to use many standard phylogenetic tools. To address these challenges, we generated a uniform reannotation of the protein-coding genes for all avian genomes based on synteny in chicken and zebra finch (SM2). We found that the SATé iterative alignment program (*53*, *54*) yielded more reliable alignments than other algorithms for large-scale data, and we developed alignment-filtering algorithms to remove unaligned and incorrectly overaligned sequences (SM3). We developed ExAML, a computationally more efficient version of the maximum likelihood program RAxML,

[1]Department of Neurobiology, Howard Hughes Medical Institute (HHMI), and Duke University Medical Center, Durham, NC 27710, USA. [2]Department of Computer Science, The University of Texas at Austin, Austin, TX 78712, USA. [3]Scientific Computing Group, Heidelberg Institute for Theoretical Studies, Heidelberg, Germany. [4]China National GeneBank, BGI-Shenzhen, Shenzhen 518083, China. [5]College of Medicine and Forensics, Xi'an Jiaotong University Xi'an 710061, China. [6]Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, 1350 Copenhagen, Denmark. [7]Department of Biology, New Mexico State University, Las Cruces, NM 88003, USA. [8]School of Biological Sciences, University of Sydney, Sydney, New South Wales 2006, Australia. [9]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90095, USA. [10]Department of Biological Sciences, Louisiana State University, Baton Rouge, LA 70803, USA. [11]CNRS UMR 5554, Institut des Sciences de l'Evolution de Montpellier, Université Montpellier II Montpellier, France. [12]Department of Evolutionary Biology, Evolutionary Biology Centre, Uppsala University, SE-752 36 Uppsala Sweden. [13]Biodiversity and Biocomplexity Unit, Okinawa Institute of Science and Technology Onna-son, Okinawa 904-0495, Japan. [14]Department of Statistics and Institute of Bioinformatics, University of Georgia, Athens, GA 30602, USA. [15]Laboratoire de Biométrie et Biologie Evolutive, Centre National de la Recherche Scientifique, Université de Lyon, F-69622 Villeurbanne, France. [16]Environmental Futures Research Institute, Griffith University, Nathan, Queensland 4111, Australia. [17]Bioinformatics and Genomics Programme, Centre for Genomic Regulation, Dr. Aiguader 88, 08003 Barcelona, Spain. [18]Universitat Pompeu Fabra, Barcelona, Spain. [19]Institució Catalana de Recerca i Estudis Avançats, Barcelona, Spain. [20]Joint Institute for Computational Sciences, The University of Tennessee, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA. [21]Bioinformatics Research Centre, Aarhus University, DK-8000 Aarhus C, Denmark. [22]The Genome Institute, Washington University School of Medicine, St Louis, MI 63108, USA. [23]Department of Biochemistry, Molecular Biology, Entomology and Plant Pathology, Mississippi State University, Mississippi State, MS 39762, USA. [24]Institute for Genomics, Biocomputing and Biotechnology, Mississippi State University, Mississippi State, MS 39762, USA. [25]Department of Biological Sciences, Texas Tech University, Lubbock, TX 79409, USA. [26]Department of Ecology and Evolutionary Biology, University of California Santa Cruz (UCSC), Santa Cruz, CA 95064, USA. [27]Organisms and Environment Division, Cardiff School of Biosciences, Cardiff University Cardiff CF10 3AX, Wales, UK. [28]Key Laboratory of Animal Ecology and Conservation Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China. [29]International Wildlife Consultants, Carmarthen SA33 5YL, Wales, UK. [30]College of Medicine and Forensics, Xi'an Jiaotong University Xi'an, 710061, China. [31]State Key Laboratory for Agrobiotechnology, China Agricultural University, Beijing 100094, China. [32]Department of Ecology and Evolutionary Biology, Tulane University, New Orleans, LA 70118, USA. [33]Museum of Natural Science and Department of Biological Sciences, Louisiana State University, Baton Rouge, LA 70803, USA. [34]Center for Zoo and Wild Animal Health, Copenhagen Zoo Roskildevej 38, DK-2000 Frederiksberg, Denmark. [35]Department of Behavioral Neuroscience, Oregon Health and Science University, Portland, OR 97239, USA. [36]Brazilian Avian Genome Consortium (CNPq/FAPESPA-SISBIO Aves), Federal University of Para, Belem, Para, Brazil. [37]Institute of Biological Sciences, Federal University of Para, Belem, Para, Brazil. [38]Institute of Medical Biochemistry Leopoldo de Meis, Federal University of Rio de Janeiro, Rio de Janeiro RJ 21941-902, Brazil. [39]Centre for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark Kemitorvet 208, 2800 Kgs Lyngby, Denmark. [40]Breeding Centre for Endangered Arabian Wildlife, Sharjah, United Arab Emirates. [41]Dubai Falcon Hospital, Dubai, United Arab Emirates. [42]Canterbury Museum Rolleston Avenue, Christchurch 8050, New Zealand. [43]Trace and Environmental DNA Laboratory Department of Environment and Agriculture, Curtin University, Perth, Western Australia 6102, Australia. [44]Department of Integrative Biology, University of California, Berkeley, CA 94720, USA. [45]Laboratory of Genomic Diversity, National Cancer Institute Frederick, MD 21702, USA. [46]Institute of Molecular and Cellular Biology, SB RAS and Novosibirsk State University, Novosibirsk, Russia. [47]Smithsonian Institution National Museum of Natural History, Washington, DC 20013, USA. [48]BGI-Shenzhen, Shenzhen 518083, China. [49]Department of Biological Sciences, National University of Singapore, Republic of Singapore. [50]Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Suitland, MD 20746, USA. [51]Center for Macroecology, Evolution and Climate, Natural History Museum of Denmark, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen Ø, Denmark. [52]Bell Museum of Natural History, University of Minnesota, Saint Paul, MN 55108, USA. [53]Department of Life Sciences, Natural History Museum, Cromwell Road, London SW7 5BD, UK. [54]Department of Life Sciences, Imperial College London, Silwood Park Campus, Ascot SL5 7PY, UK. [55]Smithsonian Conservation Biology Institute, National Zoological Park, Front Royal, VA 22630, USA. [56]Smithsonian Conservation Biology Institute, National Zoological Park, Washington, DC 20008, USA. [57]Theodosius Dobzhansky Center for Genome Bioinformatics, St. Petersburg State University, St. Petersburg, Russia 199004. [58]Oceanographic Center, Nova Southeastern University, Ft Lauderdale, FL 33004, USA. [59]Center for Biomolecular Science and Engineering, UCSC, Santa Cruz, CA 95064, USA. [60]San Diego Zoo Institute for Conservation Research, Escondido, CA 92027, USA. [61]Department of Vertebrate Zoology, MRC-116, National Museum of Natural History, Smithsonian Institution, Washington, DC 20013, USA. [62]Department of Environmental Health Science, University of Georgia, Athens, GA 30602, USA. [63]Moore Laboratory of Zoology and Department of Biology, Occidental College, Los Angeles, CA 90041, USA. [64]Department of Genomics and Genetics, The Roslin Institute and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush Campus, Midlothian EH25 9RG, UK. [65]Swedish Species Information Centre, Swedish University of Agricultural Sciences Box 7007, SE-750 07 Uppsala, Sweden. [66]Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China. [67]Department of Organismic and Evolutionary Biology and Museum of Comparative Zoology, Harvard University, Cambridge, MA 02138, USA. [68]Institute of Theoretical Informatics, Department of Informatics, Karlsruhe Institute of Technology, D- 76131 Karlsruhe, Germany. [69]Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94158, USA. [70]Department of Ornithology, American Museum of Natural History, New York, NY 10024, USA. [71]Department of Biology and Genetics Institute, University of Florida, Gainesville, FL 32611, USA. [72]Departments of Bioengineering and Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA. [73]Department of Biology, University of Copenhagen, Ole Maaløes Vej 5, 2200 Copenhagen, Denmark. [74]Princess Al Jawhara Center of Excellence in the Research of Hereditary Disorders, King Abdulaziz University, Jeddah 21589, Saudi Arabia. [75]Macau University of Science and Technology, Avenida Wai long, Taipa, Macau 999078, China. [76]Department of Medicine, University of Hong Kong, Hong Kong. [77]Centre for Social Evolution, Department of Biology, Universitetsparken 15, University of Copenhagen, DK-2100 Copenhagen, Denmark.
*These authors contributed equally to this work. †**Corresponding author. E-mail: jarvis@neuro.duke.edu (E.D.J.); tandywarnow@gmail.com (T.W.); mtpgilbert@gmail.com (M.T.P.G.); wangj@genomics.cn (W.J.); zhanggj@genomics.cn (G. Z.)**

for estimating species trees from genome-scale concatenated sequence alignments (SM4) (*55*–*57*). We also developed a statistical binning approach that improves multispecies coalescent analyses for handling gene trees with low phylogenetic signal to infer a species tree (SM5) (*58*). These computationally intensive analyses were conducted on more than 9 supercomputer centers and required the equivalent of >400 years of computing using a single processor (SM3 and SM4).

From these efforts, we identified a high-quality orthologous gene set across avian species, consisting of exons from 8251 syntenic protein-coding genes (~40% of the proteome), introns from 2516 of these genes, and a nonoverlapping set of 3769 ultraconserved elements (UCEs) with ~1000 bp of flanking sequences. This total evidence nucleotide data set comprised ~41.8 million bp (table S3 and SM4), representing ~3.5% of an average avian genome.

A genome-scale avian phylogeny

Total evidence nucleotide tree

The total evidence nucleotide alignment partitioned by data type (introns, UCEs, and first and second exon codon positions; third positions excluded as described later) analyzed with ExaML under the GTR+GAMMA model of sequence evolution (SM4) resulted in a highly resolved total evidence nucleotide tree (TENT) (Fig. 1 and fig. S1). The three recognized major groupings within extant birds—Palaeognathae, Galloanseres, and Neoaves (the latter two united in the infraclass Neognathae)—were recovered with full (100%) bootstrap support (BS). The tree revealed the first divergence within extant Neoaves, resulting in two fully supported, reciprocally monophyletic sister clades that we named Passerea (after its most speciose group Passeriformes) and Columbea (after its most speciose group Columbiformes) (Fig. 1; see SM6 for rationale of clade names).

Within Passerea, the TENT strongly confirmed the monophyly of two large closely related clades that we refer to as core landbirds (Telluraves) and core waterbirds (Aequornithia) (*8*, *16*, *17*, *27*, *36*, *59*);

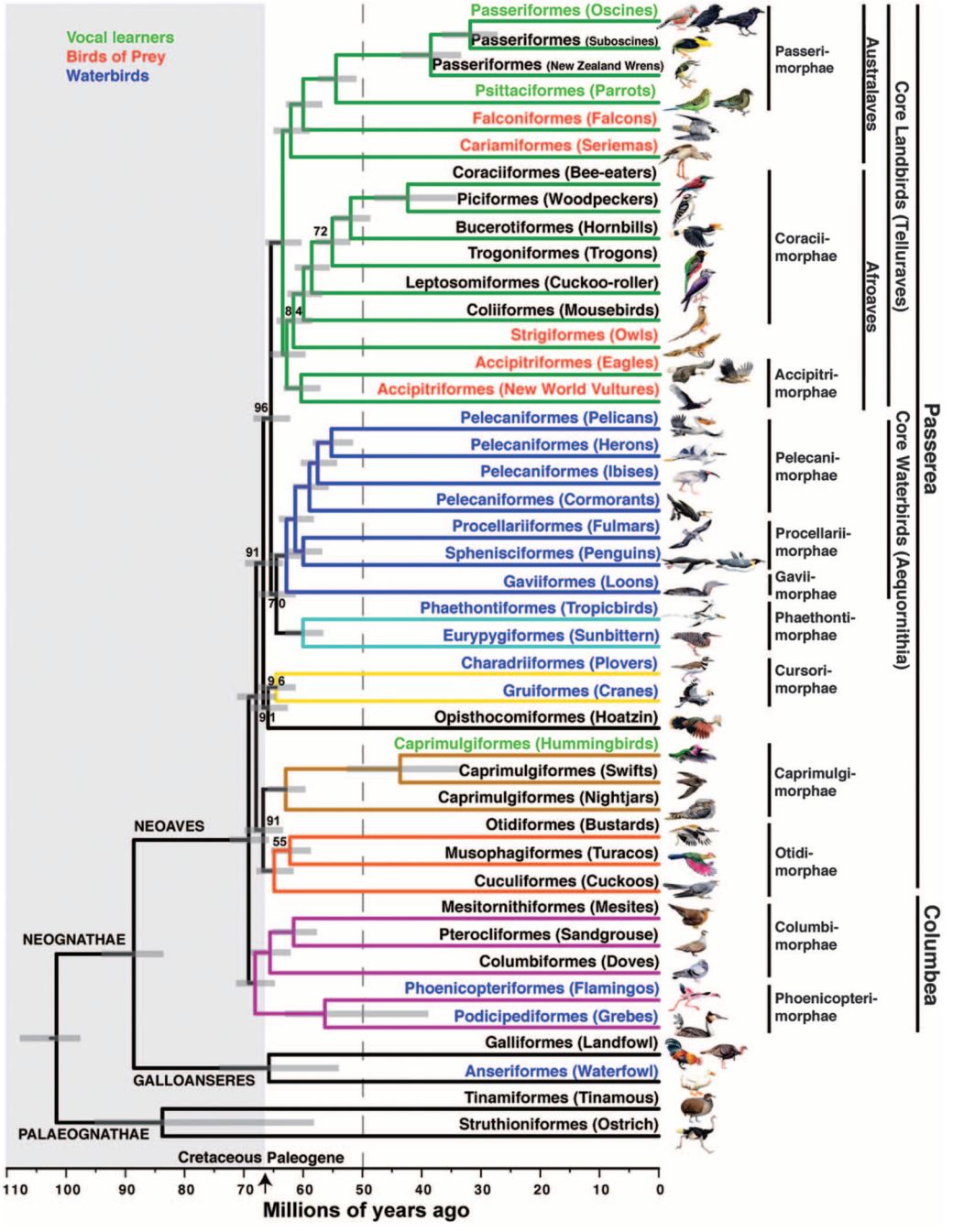


Fig. 1. Genome-scale phylogeny of birds. The dated TENT inferred with ExaML. Branch colors denote well-supported clades in this and other analyses. All BS values are 100% except where noted. Names on branches denote orders (-iformes) and English group terms (in parentheses); drawings are of the specific species sequenced (names in table S1 and fig. S1). Order names are according to (*36*, *37*) (SM6). To the right are superorder (-imorphae) and higher unranked names. In some groups, more than one species was sequenced, and these branches have been collapsed (noncollapsed version in fig. S1). Text color denotes groups of species with broadly shared traits, whether by homology or convergence. The arrow indicates the K-Pg boundary at 66 Ma, with the Cretaceous period shaded at left. The gray dashed line represents the approximate end time (50 Ma) by which nearly all neoavian orders diverged. Horizontal gray bars on each node indicate the 95% credible interval of divergence time in millions of years.

we use the term "core" instead of "higher" to prevent interpretation that these groups are more advanced or more recently evolved than other birds. Within core landbirds, we found 100% BS for a previously more weakly supported clade (Australaves) containing seriemas (historically placed in Gruiformes), falcons (historically grouped with other diurnal birds of prey), parrots (historically difficult to place), and Passeriformes and a sister clade (Afroaves) containing Acciptrimorphae birds of prey, owls, mousebirds, woodpeckers, and bee eaters, among others (Fig. 1) (*8*, *17*, *26*, *29*, *60*). Core waterbirds were sister to a fully supported clade (Phaethontimorphae) containing tropicbirds and the sunbittern (Fig. 1) (*27*, *28*). We did not include Phaethontimorphae in the core waterbirds because their relationship had relatively low 70% BS, although their aquatic (tropicbirds)

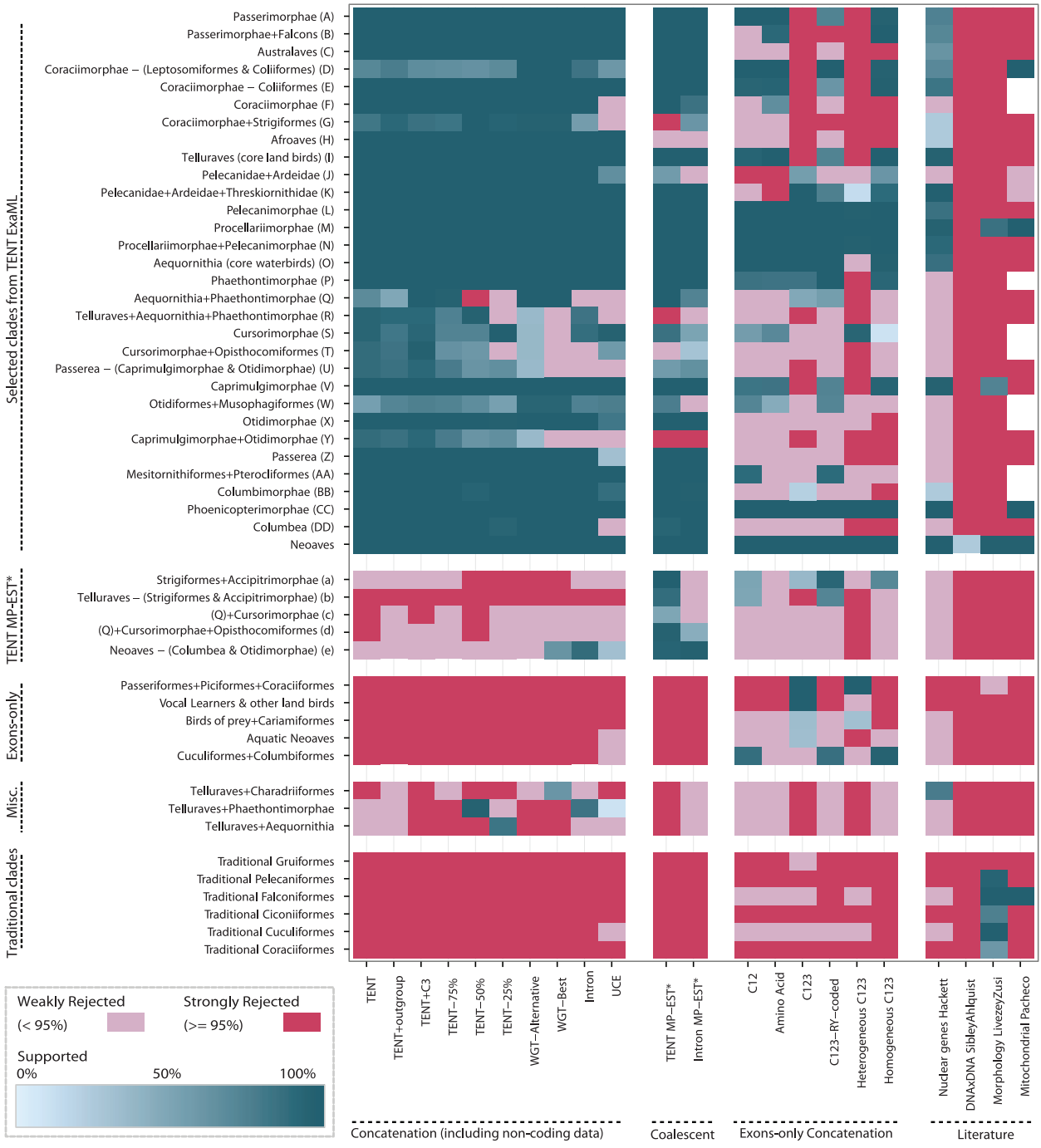


Fig. 2. Metatable analysis of species trees. Results for different genomic partitions, methods, and data types are consistent with or contradict clades in our TENT ExaML, TENT MP-EST*, and exon-only trees and previous studies of morphology (*15*), DNA-DNA hybridization (*24*), mitochondrial genes (*14*), and nuclear genes (*17*). Letters (A to DD and a to e) denote clade nodes highlighted in Fig. 3, A and B, of the ExaML and MP-EST* TENT trees. Each column represents a species tree; each row represents a potential clade. Blue-green signifies the monophyly of a clade, and shades show the level of its BS (0 to 100%). Red, rejection of a clade; white, missing data. We used a 95% cut-off (instead of a standard 75%) for strong rejection due to higher support values with genome-scale data. The threshold for the mitochondrial study was set to 99% due to Bayesian posterior probabilities yielding higher values than BS. An expanded metatable showing partitioned ExaML, unbinned MP-EST, and additional codon tree analyses is shown in fig. S2.

and semiaquatic (sunbittern) lifestyles are consistent with a waterbird grouping, and multiple analyses presented below group them with 100% BS. The TENT also resolved at 100% BS taxa that were previously difficult to place, including uniting cuckoos, turacos, and bustards (Otidimorphae) and placement of the mousebird among core landbirds. The Columbea also had separate landbird and waterbird groups. These results demonstrate that genome-scale data can help resolve difficult relationships in the tree of life.

Comparisons of TENT with previous studies

The TENT contradicted some relationships in avian phylogenies generated from morphological characters (*15*), DNA-DNA hybridization (*24*), and mitochondrial genomes (*14*, *18*) (Figs. 2, fig. S2, and Fig. 3A versus fig. S3, A to C). For example, our Falconiformes excluded the previously included eagles and New World vultures (now in Acciptrimorphae); our Coraciiformes was more narrowly delineated and excluded hornbills and cuckoo-rollers; our Pelecaniformes excluded tropicbirds; and our Gruiformes excluded seriemas, bustards, the sunbittern, and mesites. The TENT did not fully support the view based on one gene (*β-fibrinogen*) that the first divergence in Neoaves resulted in two equally large Metaves and Coronaves radiations (*25*). However, all Columbea species in the TENT were in the previously defined Metaves, supporting the hypothesis of two parallel radiations of birds with convergent adaptations (*25*).

The TENT was most congruent with past (*8*, *17*) and more recent (*27*, *28*) smaller-scale multilocus nuclear trees (Figs. 2 and 3A and fig. S3D), although most congruence was limited to the core landbirds and core waterbirds. Within the former, we recovered Australaves and Afroaves (*60*), although with a different branching ordering in our tree; our taxon sampling is insufficient to address the biogeographic justification of their names. The TENT recovered a number of groups not present in these previous trees, and even for those present, the TENT had higher BS (Fig. 2). Absence of nonavian outgroups in our TENT above was not responsible for variation with past studies because we recovered the same topology when including outgroups (Fig. 2 and fig. S4, A and B), despite the outgroups having only ~30% orthologous sequences in the TENT alignment (e.g., fig. S21; SM3).

More data are responsible for resolving early branches of the tree

Despite the many fully supported (100% BS) relationships in the TENT, lower support was obtained for 9 of the 45 internal branches (although still within the high 70 to 96% BS range). Almost all were at deep divergences within the Neoaves, after the Columbea and Passerea divergence and before the ordinal divergences (Fig. 1 and fig. S1). The monophyly of each of the superorders, however, had 100% BS. The presence of these lower BS values is in contrast to the expectation that genome-scale alignments would result in complete phylogenetic resolution (*34*, *35*, *61*).

However, consistent with this hypothesis, we found that most relationships that had less than 100% BS with the full TENT data exhibited a steady increase in support with an increase in random subsets of the TENT data (Fig. 2 and fig. S5). The placement of the Phaethontimorphae (sunbittern and tropicbirds) and hoatzin changed when smaller (25 to 50%) amounts of data were analyzed. Further exploring data amount, we used the assembled ~1.1-billion-bp chicken genome (*40*) as a reference to generate a 322-million-bp MULTIZ alignment of putatively orthologous genome regions across all species, comprising ~30% of an average assembled avian genome and corresponding to the maximal orthologous sequence obtainable across all orders under our homology criteria (SM3). We ran ExaML on the alignment for ~42 CPU years, with 20 maximum likelihood searches on distinct starting trees and 50 bootstrap replicates before reaching our convergence criterion (SM4) on a whole-genome tree (WGT). Notably, all runs resulted in one of two trees: one identical to the TENT topology (fig. S4C) and a second almost identical to the TENT (fig. S4D). This latter tree differed from the TENT by local shifts in five branches, all clades that had less than 100% BS in the TENT (fig. S4, A and D). Given the relatively minor differences between the second WGT and the TENT, together they corroborate the majority of relationships in the avian tree of life. Although the WGT has more data (table S3), the orthology (SM2) and alignment (SM3) qualities are higher for the TENT, and thus we consider the TENT more reliable.

Noncoding data contribute more to the TENT topology

We sought to determine if different genomic partitions contribute differently to the TENT and found that ExaML trees using only introns or UCEs from the TENT data were largely congruent with the TENT and WGTs for branches that had strong support (BS > 75%) in the intron and UCE trees (Figs. 2; 4, A and B; and 5B). However, the intron tree, and even more so the UCE tree, had lower resolution than the TENT (Fig. 5A), mostly on deep branches (Fig. 4, A and B), consistent with fewer data leading to lower resolution on deeper branches. For the intron tree, some lower-resolution branches had local shifts, but they matched those found in the second WGT or the 25 to 75% data subsets of the TENT; an exception was Phaethontimorphae, which moved from being sister to core waterbirds with 70% BS in the TENT (but 100% BS in the WGTs) to sister to core landbirds with 86% BS in the intron tree. For the UCE tree, the lower-resolution, deep branches had more distant shifts. Trees created from analysis of the first and second codon positions (exon c12) of the TENT data also had lower levels of BS (~39 to 64%) but with more topological differences on the deep branches (Figs. 2, 4C, and 5A), yet all but one of the fully resolved relationships (local difference in egret + ibis + pelican) were congruent with the TENT (Fig. 5B).

These findings demonstrate that noncoding intron sequences lend greater support for the TENT than the protein-coding and UCE sequences, consistent with intron sequences having a higher rate of evolution (SM4) and thus greater phylogenetic signal. These differences are not merely due to shorter alignments of the exon and UCE sequences, because each accounted for ~25% of the TENT data, similar in sequence length to the random 25% subset of the TENT with introns (table S3) that produced a tree with a higher average BS and a topology closer to the full TENT (Fig. 5A and fig. S5D).

Incomplete lineage sorting and impact on deep branches

Deeper branches exhibit higher gene tree incongruence

We next investigated ILS, a population-level process that results in incongruence between gene trees and the species tree (*62*). Consistent with conditions that could lead to ILS (*63*), the TENT had a wall of many (25 of 45; ~55%) very short internal branches (0.0006 to 0.002 substitutions per site), almost all at deep divergences within Neoaves (Fig. 3A, inset, and fig. S7). Indeed, all nine branches with <100% BS were among the shortest in the TENT (fig. S8), many in succession, suggesting that reduced BS could be related to conflict among gene trees.

To test this hypothesis, we compared the distribution of gene trees that have strong conflict (>75% BS) with branches of the ExaML TENT. We focused on introns because they had greater gene tree resolution (higher average BS) than exons or UCEs (fig. S24 and SM4). The 2485 introns with orthologs available in the two outgroup Palaeognathae species ranged from exhibiting no conflict to exhibiting considerable conflict (up to 950 genes or 38%) for some branches of the TENT (Fig. 3A, blue numbers, and Fig. 5C). The percentage of gene tree conflict was successively higher for the shorter and deeper branches of the TENT (Fig. 3A), particularly those with <100% BS (e.g., branches R, U, and Z in Fig. 3, A and C). Conversely, these short branches had fewer (0 to 20%) intron gene trees supporting them at high (>75%) BS levels (Fig. 3A, black numbers, and Fig. 3D). These findings suggest that ILS could have affected the inferred relationships of some of the deep branches of Neoaves in the concatenated tree analysis.

Multispecies coalescent approach infers a species tree similar to the TENT

To determine if ILS affected the concatenated tree analyses, we explored whether a multispecies coalescent model leads to a different tree topology. Multispecies coalescent methods estimate the species tree from a set of gene trees and are statistically consistent when discordance among gene trees results from ILS (*64*, *65*). However, the inferred species tree can have low resolution (BS) and be less topologically accurate when the input gene trees are poorly resolved (*33*, *66*), a problem that many of our genes faced (SM4). Thus, we developed a statistical binning technique that first groups genes into sets based on phylogenetic similarities, from each set estimates a supergene tree, and uses them in the maximum pseudolikelihood estimation of the species tree

(MP-EST) multispecies coalescent approach (*67*) to infer a species tree (SM5) (*58*).

This approach produced more accurate estimated species trees compared with MP-EST applied to unbinned gene data sets that have low phylogenetic signal (i.e., figs. S2 and S9; SM7) (*58*). It produced a highly resolved binned MP-EST (MP-EST*) TENT tree that was highly congruent with the ExaML TENT (Fig. 3, A and B). There were only local shifts of five clades, nearly all on lower-support (<100% BS) branches of the ExaML and MP-EST* TENTs (Fig. 3, A and B). The monophyly of Afroaves was the only case of 100% BS in the ExaML TENT that conflicted with the MP-EST* TENT tree and involved a local shift in

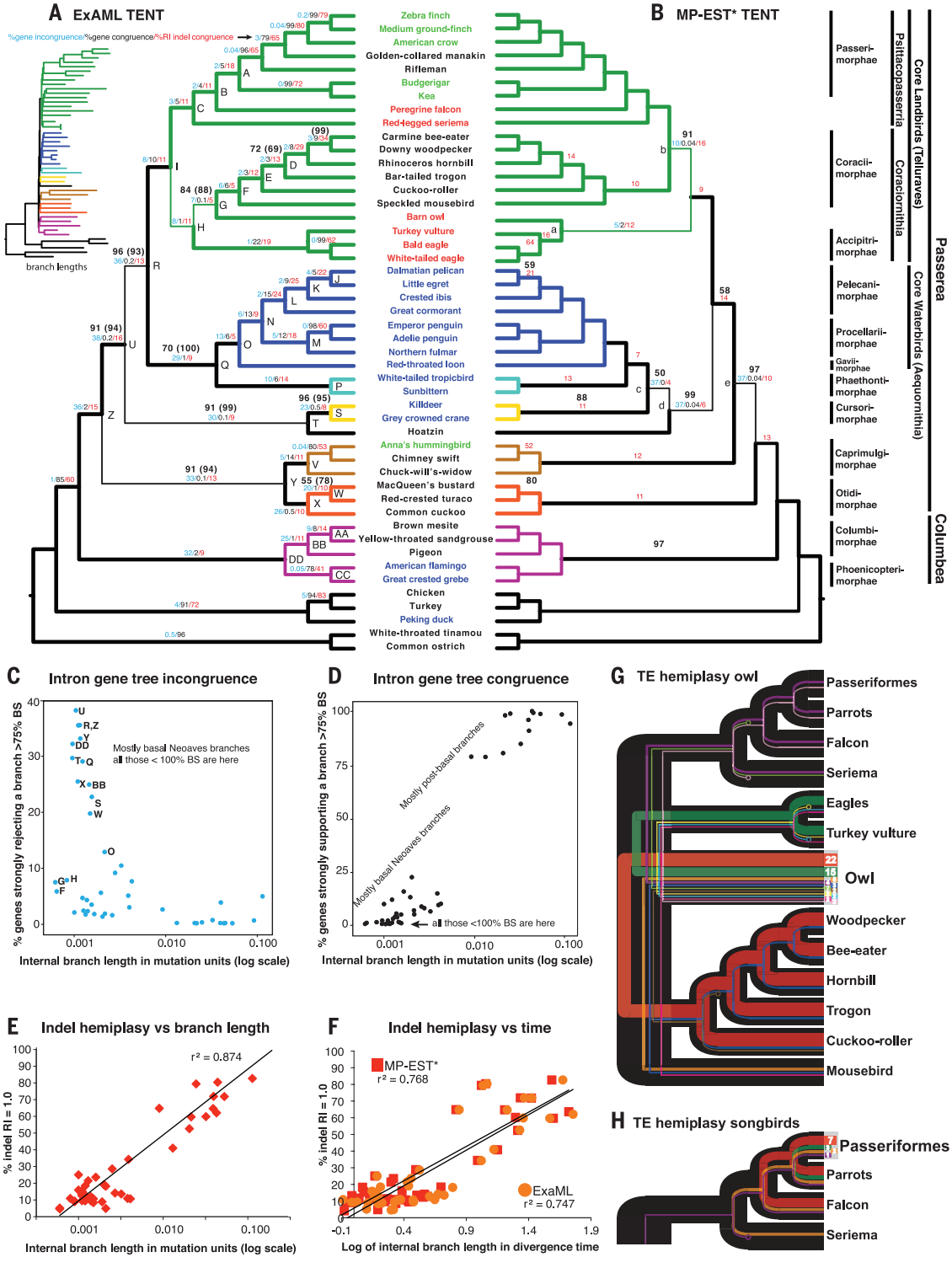


Fig. 3. Evidence of ILS. (**A**) Cladogram of ExaML TENT avian species tree, annotated for nodes from Fig. 2 (letters), for branches with less than 100% BS without and with (parentheses) third codon positions, for strong (>75% BS) intron gene tree incongruence and congruence, and for indel congruence on all branches (except the root). Thin branch lines represent those not present in the MP-EST* TENT of (B). (Inset) ExaML branch lengths in substitution units (expanded view in fig. S7). Color coding of branches and species is as in Fig. 1. (**B**) Cladogram of MP-EST* TENT species tree, annotated similarly as in the ExaML TENT in (A). Thin branch lines represent those not present in the ExaML TENT of (A). (**C**) Percent of intron gene trees rejecting (≥75% BS) branches in the ExaML TENT species tree relative to branch lengths. Letters denote nodes in (A) that either have less than 100% support or are different from the MP-EST* TENT in (B). (**D**) Percent of intron gene trees supporting (≥75% BS) branches in the ExaML TENT species tree relative to branch lengths. (**E**) Indel hemiplasy [the inverse of percent of retention index (RI) = 1.0 indels that support the branch; see SM9] correlated with ExaML TENT branch length (log transformed). r^2, correlation coefficient. (**F**) Indel hemiplasy correlated with ExaML and MP-EST TENT internal branch divergence times in millions of years (log transformed). Plotting with internal branch times was necessary to compare both trees (SM9). (**G**) TE hemiplasy with owls among the core landbirds. Line color, shared TE tree topology; line thickness, relative proportion of TEs that support a specific topology (total numbers shown in the owl node). Circles at end of lines indicate loss of the TE allele in that species after ILS, as the sequence assembly contains an empty TE insertion site (SM10). Only topologies with two or more TEs are shown. (**H**) TE hemiplasy with songbirds among the core landbirds.

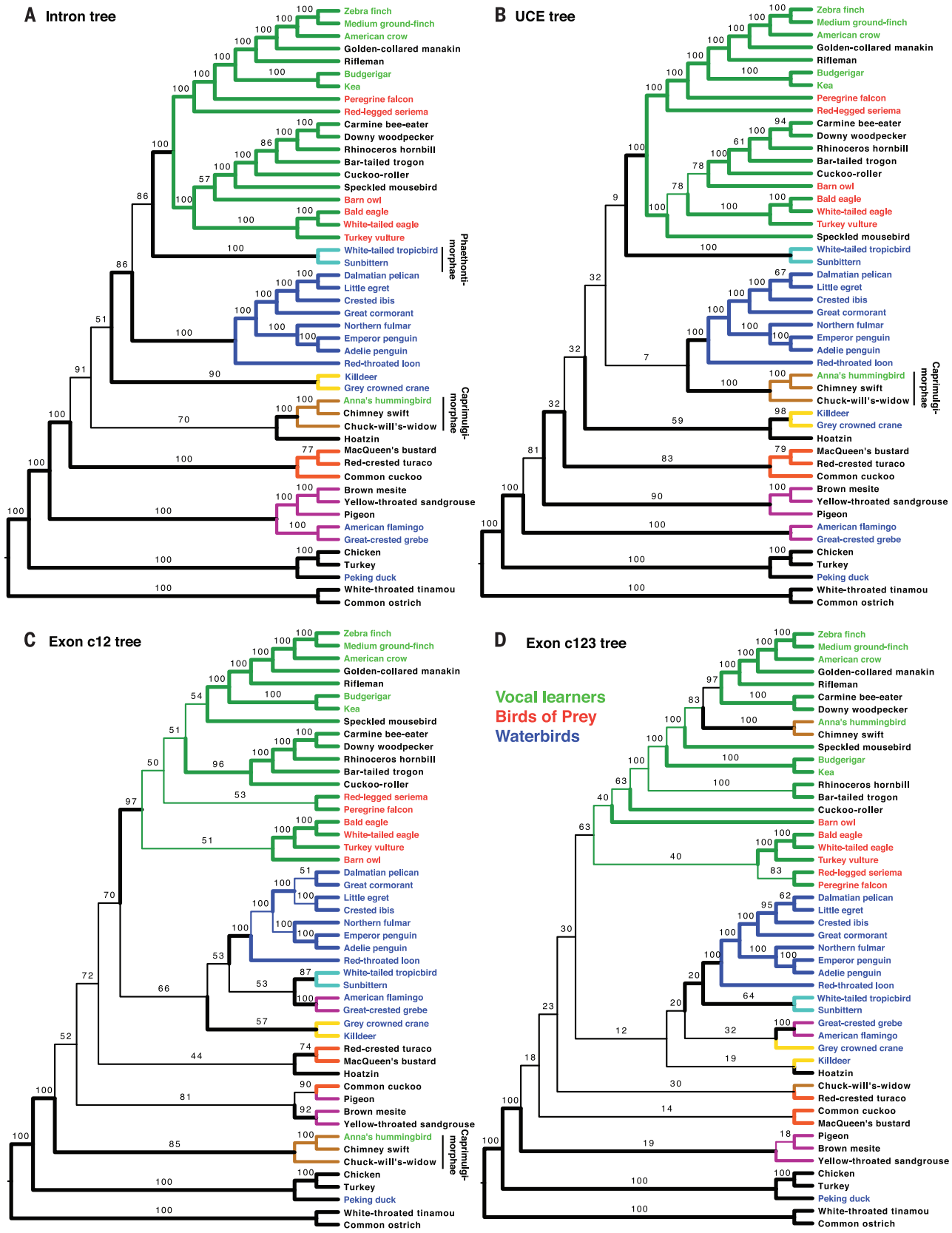


Fig. 4. Species trees inferred from concatenation of different genomic partitions. (**A**) Intron tree. (**B**) UCE tree. (**C**) Exon c12 tree. (**D**) Exon c123 tree. The tree with the highest likelihood for each ExaML analysis is shown. Color coding of branches and species is as in Fig. 1 and fig. S1. Thick branches denote those present in the ExaML TENT. Numbers give the percent of BS.

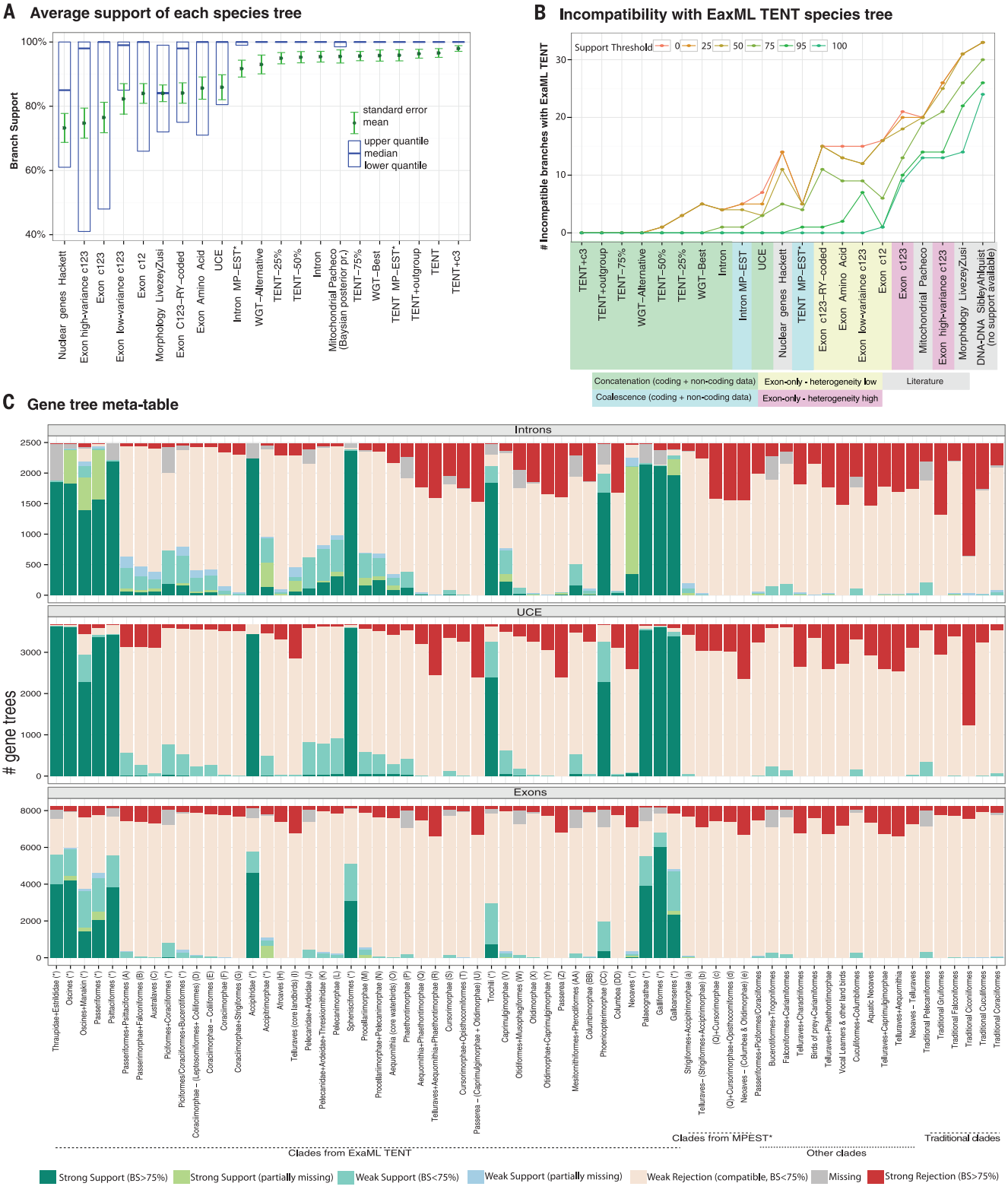


Fig. 5. Comparisons of total support among species trees and gene trees. (**A**) Average BS across all branches of species trees from varying input data as in Fig. 2, ordered left to right from lowest to highest values. (**B**) Numbers of incompatible branches (out of 45 internal), at different support thresholds, with the ExaML TENT tree, ordered left to right from most to least compatible (expanded analysis in fig. S6). (**C**) Analyses of intron, exon, and UCE gene tree congruence and incongruence with nodes in the ExaML TENT, MP-EST* TENT, and other species trees. Names and letters for clades are as in Figs. 2 and 3. "Missing" denotes the case in which an ortholog is not present for any taxa or is present for only one taxon, and hence monophyly cannot be determined. "Partially missing" indicates the case in which some taxa are missing but at least two of the taxa are present, and thus we can still categorize it as either monophyletic or not. For further details, see SM7.

the owl with mousebirds and Acciptrimorphae birds of prey. Two branches with <100% BS in the ExaML TENT increased to 100% in the MP-EST* TENT, including Phaethontimorphae with core waterbirds. The intron trees supported some branches more in the ExaML and some more in the MP-EST* TENT (Fig. 3, A and B). Nevertheless, the overall topology of both trees was very similar, including the basal Columbea and Passerea divergence.

All estimates of gene trees differ from our candidate species trees

No single intron, exon, or UCE locus from our TENT data set had an estimated topology identical to the ExaML TENT or MP-EST* TENT (fig. S10, A and B). The top three loci (all introns) with the closest inferred topologies differed from the two versions of the TENT on more than 20 to 30% of their branches. Average topological distance with the ExaML species tree was 63% for the introns, 66% for the UCEs, and 80% for the exons. To test whether our total evidence data set missed some genes with the TENT topologies, we constructed a more comprehensive collection of genes trees with phylomeDB, which assigns orthology using maximum likelihood analyses (http://phylomedb.org) [see SM8 and (*68*)]. For ~13,000 (low-coverage genomes) to ~18,000 (high-coverage genomes) annotated genes across avian species (*44*), phylomeDB inferred orthologs for 94.58% of them and these agreed with the synteny-based orthology of the 8251 protein-coding genes of the TENT by 93%. This more complete set of protein-coding genes still did not have a single estimated gene tree that was fully congruent with the ExaML or MP-EST* TENT trees (fig. S10, C and D), and there was overall low congruence with the species trees (http://tol.cgenomics.org/birds_v1) (fig. S11, A and B). The conflicting nodes largely reflected branches with low statistical support (approximate likelihood ratio test < 0.95), which primarily corresponded to the short successive deep branches of Neoaves. These findings can be explained by both a low amount of phylogenetic signal in individual loci (figs. S24 to S26 and SM4) and a high amount of ILS during the neoavian radiation.

Indels suggest a high degree of ILS at the earliest branches of the Neoaves tree

We further assessed ILS using insertions and deletions (indels) (*69*), because they have less homoplasy (convergence) than single nucleotides (SM9), and unlike gene trees, indels can be examined as discrete characters mapped to a reference tree without the added inference of constructing trees from them. We scored 5.7 million indels from the TENT alignment, of which 24% were shared by two or more taxa (table S3). We found indel incongruence on all branches of the ExaML TENT, as measured inversely by the percent of the indel characters uniquely defining each branch (Fig. 3A, red numbers; SM9). Like the gene trees, there appeared to be a successive decrease in the percentage of indels that supported deeper branches of each major clade (Fig. 3A). Most branches with the highest levels of indel incongruence belonged to the shortest and deepest ones that made local shifts in analyses, with the two branches joining mousebirds and owls exhibiting the highest indel incongruence and the shortest internal branch lengths in the ExaML TENT (Fig. 3A and fig. S7). Consistent with these findings, indel incongruence was inversely correlated with internal branch length, and branch length explained 87% (r^2) of the variation in the percentage of nonhomoplasious indels defining each branch (Fig. 3E). The correlation of indel incongruence versus branch time was similar for both ExaML and MP-EST* TENT trees (Fig. 3F).

Indel incongruence is not due to the indels supporting another species tree, as applying ExaML on indels from the total evidence alignment as binary data produced a total evidence indel tree that was largely congruent with the ExaML TENT and MP-EST* TENT for all but one node with a local shift of pigeon within Columbea (fig. S12). Homoplasy due to convergence is thought to be positively correlated with branch length [i.e., long branch attraction (*70*)]. The only known source of incongruence that is inversely correlated with internal branch length is hemiplasy (differential inheritance of polymorphic alleles) (*64*, *71*). Because hemiplasy is a hallmark of ILS and 87% of the variation in indel incongruence is explained by branch length, our indel findings suggest high levels of ILS during the basal radiation of Neoaves, with comparable support for the ExaML or MP-EST* versions of the TENT.

Transposable elements with higher ILS in the deepest branch of core landbirds with owls

We tested for a signature of ILS in TE insertions, which have extremely low homoplasy because independent insertions into the same location in a genome are rare (SM10) (*72*, *73*). We focused on the owl because its position exhibited one of the strongest incongruencies among the species tree results. Of 3671 barn owl long terminal repeat TE insertion loci orthologous in all of the bird genomes, 61 were informative for owls among core landbirds and showed two dominant exclusive TE topologies: (i) an owl + Accipitrimorphae topology, as seen in the MP-EST* TENT; and (ii) an owl + Coraciimorphae topology that excludes mousebird, as seen in the UCE tree (Fig. 3G compared to Figs. 3B and 4B). Nine other topologies had fewer markers supporting them. In contrast, for 25 informative TEs of Neoaves in (*29*), 13 were informative for Australaves, and of these, 3 were exclusive for Passeriformes + parrots, 7 for Passeriformes + parrots + falcons, and 2 for the latter group plus seriemas, with no alternative topologies for the first two groups (Fig. 3H). If the passeriform TE insertions exhibited a similar mixture of alternative distributions as for the owl, just 10 markers would result in conflicting distributions (4 with one, 3 with another, and 3 for the remaining topologies) instead of a conflict-free topology. Although this analysis is limited to specific taxa, it suggests higher ILS near the deepest branches of Afroaves involving the owl, consistent with the branch length, gene tree, and indel findings.

Overall, these results reveal considerable ILS during the neoavian radiation and that, even with genome-scale data, ILS may affect the inference of small local relationships in the deep branches of the species tree that have long been more challenging to resolve. However, ILS does not affect the majority of other phylogenetic relationships we found using genome-scale data.

Protein-coding data resolve avian phylogeny poorly but reflect life history traits

Codon positions of protein-coding genes and life history relationships

We investigated sources of lower resolution and incongruence for the tree based on protein-coding sequences (Fig. 4C). This is crucial for phylogenomic inference, as many studies [including transcriptome analyses (*19*, *74*)] use only protein-coding genes to infer species trees. We found that ExaML analyses with either all (c123; Fig. 4D) or individual codon positions (c1, c2, c3; fig. S13, A to C) produced trees with lower BS (Fig. 5A) and greater differences in topologies (Fig. 2 and fig. S2) compared with noncoding data and coding + noncoding combined. The differences between coding versus noncoding trees were not solely due to shorter sequence length of the coding data, because the full coding data set (13.3 million bp for c123) produced a tree with fully supported (100% BS) relationships that were incongruent with those fully supported in the intron (19.3 million bp), TENT (37.4 million bp without the third codon position), and WGT (322.1 million bp) (Figs. 2 and 5B, and table S3). Surprisingly, the c123 topology associated species more with life history traits than the TENT topology. This included a strongly supported clade (100% BS on most branches) that comprised the three groups of vocal learners (parrots, songbirds, and hummingbirds) and most of the nonpredatory core landbirds, a monophyletic clade of diurnal birds of prey and seriemas (albeit with low 40% BS), and a monophyletic clade of all aquatic and semiaquatic species of Passerea and Columbea (also with low 20% BS) (Fig. 4D). Partitioning the data to account for possible differences in evolutionary rates among genes (SM4) did not result in a tree more similar to the TENT, but instead in a tree with increased support for monophyletic groupings of species with these broadly shared traits (fig. S14C). The c1, c2, and amino acid tree topologies (fig. S13, A, B, and D) were more congruent with the c12 tree (Figs. 2 and 4C), consistent with these two codon positions largely specifying amino acid identity. In contrast, the c3 tree was very similar to the c123 tree but with higher BS (63 to 82%) for similar trait groupings; it moreover brought all basal neoavian landbirds together as sister to all neoavian aquatic/semiaquatic species (figs. S2 and S13C). Most individual gene trees show weak to strong rejection of these relationships (Fig. 5C).

As expected (*19*), the third codon position exhibited greater base composition variation among species than the other codon positions and even other genomic partitions (fig. S15A). Although all codon positions violated the stationarity assumption

in the GTR + GAMMA model of sequence evolution, the third codon position exhibited a much stronger violation (fig. S15B). Reducing this variation by RY recoding of purines (R) and pyrimidines (Y) on the third codon position (SM4) made the c123 tree topology more similar to the c12 topology (Fig. 2 and fig. S14D). These results demonstrate that the third codon position exerts a strong influence on the protein-coding–tree topology, overriding signals from the first and second codon positions. They also suggest that a signal in the third codon position could also be associated with convergent life history traits.

Heterogeneous protein-coding genes associated with life history traits

We further investigated the source of the conflict in the protein-coding genes (SM11) and found that trees using all codon positions from the 10% most compositionally homogeneous (low-variance) exons (n = 830) were most congruent with the c12 tree and, thus, more similar to the TENT than to the c123 tree (Figs. 2 and 6A; cladograms in fig. S16, A to C). Conversely, trees using all codon positions from the 10% most compositionally heterogeneous (high-variance) genes (n = 830) were more congruent with the exon c123 and c3 trees (Figs. 2 and 6B and fig. S16, B and D). The branch lengths of the high-variance exon tree showed a strong positive correlation with GC content and a negative correlation with the average body mass of species, seen at a much lesser magnitude in the low-variance exon tree (Fig. 6, A to D). The correlations for the high-variance genes were also strongest on the third codon position (fig. S17, A and B) (*75*, *76*). In addition, the genomic positions of the high-variance genes were skewed toward the ends of the chromosomes, whereas the positions of the low-variance genes were skewed toward the center (Fig. 6, E and F, and fig. S17, C and D). Although the available introns of these genes had significant correlations among GC content and body mass and among GC content and chromosome position, they exhibited less heterogeneity overall (fig. S17, A to D) and yielded trees that were much more congruent with each other and with the TENT (figs. S2 and S17, E and F). An ExaML TENT tree that included the third codon position (TENT + c3) was identical in topology to the ExaML TENT without the third codon position and had increased support for six of the nine branches that had less than 100% BS (fig. S1 versus fig. S18, also Figs. 3A and 5A).

These results suggest that in the context of protein-coding data only, high–base compositional heterogeneity and life history have a strong impact on incongruence with the species tree, and thus are not suitable for generating a highly resolved phylogeny. However, in the context of large amounts of noncoding genomic data, the phylogenomic signal in the exon data adds support to the species tree.

Dating the radiation of Neoaves

The generation of a well-resolved avian phylogeny allowed us to address the timing of avian diversification. To estimate the avian timetree with genomic-scale data, we used first and second codon positions from 1156 clock-like exon genes (which do not strongly exhibit the above protein-coding compositional bias), calibrated with 19 conservatively chosen avian fossils (plus nonavian outgroups) as minimum bounds for lineage ages (with a maximum-bound age constraint of 99.6 Ma for Neornithes), in a Bayesian autocorrelated relaxed clock method using MCMCTREE (*77*) on the fixed ExaML TENT topology (SM12).

Our results suggest that after the Palaeognathae and Neognathae divergence about 100 Ma in the Late Cretaceous, the Palaeognathae diverged into their two stem lineages [ratites and tinamous (*11*, *78*)] about 84 Ma, and the Neognathae diverged into their stem lineages (Galloanseres and Neoaves) about 88 Ma (Fig. 1). Although the 95% credibility interval for the ostrich-tinamou divergence is broad, its lower bound is consistent with the fossil record (*79*). In contrast, both the earliest divergence within Galloanseres and an explosive diversification within Neoaves were dated to occur around the K-Pg boundary, with 95% credibility intervals spanning 6.5 million years, on average. In particular, the most basal divergences within Neoaves (Columbea, Passerea, and two more) occurred before the K-Pg transition (67 to 69 Ma) and all others after, with nearly all ordinal divergences completed by 50 Ma (Fig. 1, dashed line). The estimated age for the basal split of Passeriformes, representing ~60% of all living ~10,400 avian species, was around 39 Ma. These divergence times conflict with some previous studies based on nuclear (*9*–*12*) and mitochondrial (*13*, *14*) DNA but are consistent with the fossil record (*80*), including the identification of *Vegavis iaai*, a very Late Cretaceous (66 to 68 Ma) stem-anseriform fossil (*80*, *81*), and the dearth of verifiable Neoaves fossils in the Late Cretaceous (*5*). These findings were similar regardless of the specific tree from this study we dated or whether we used a later minimum age (86.5 Ma) for Neornithes (table S16; more discussion on dating in SM12).

Discussion

Our study is an example of the extraordinary amount of genomic sequence data required to produce a highly supported phylogeny spanning a rapid radiation. The conflict we observe with other data types (*14*, *15*, *24*) can no longer be considered to be due to error from smaller amounts of sequence data (*8*, *17*) nor to differences in concatenation versus coalescence methods (*27*, *28*). The absence of a single gene tree identical to the avian species tree is consistent with studies in yeast (*82*), indicating that phylogenetic studies based on one or several genes, especially for rapid radiations, will probably be insufficient. The major sources of the gene tree incongruence we find are low-resolution gene trees and substantial ILS during the rapid radiation. It is possible that some of the deep branches of the species tree are in the anomaly zone (*63*), although the gene tree support is not high enough to confidently test this idea. It is also possible that some gene and local species tree incongruence could reflect ancient hybridization during the radiation, but distinguishing between this and other sources of hemiplasy (*83*) would require more complete assemblies, genes without missing data across species, and development of new methods (*84*). Finally, it is also possible that insufficient taxon sampling contributed to the local species tree incongruence, known to lead to long-branch attraction (*70*). We did seek to break up some long branches, specifically within core landbirds and core waterbirds. However, the very large-scale data collection for this study made it necessary to prioritize species for specific parts of the tree. Moreover, the potential to add taxa that will break up long branches is limited for a number of groups because the species either are extinct or there are no more major lineages to sample, suggesting that further study of analytical methods for whole genomes will prove to be as important as additional taxa.

Genomic-scale amounts of protein-coding sequence data were not only insufficient but were also misleading for generating an accurate avian phylogeny due to convergence. One possible explanation is convergent GC-biased gene conversion in exons, where AT-GC mismatches are corrected by DNA repair molecules in a biased manner to produce more gametes with the GC allele (*85*). GC-biased gene conversion correlates with recombination rate (*86*), and new GC alleles reach fixation more easily in species with larger population sizes, which tend to also have smaller body sizes (*87*). Recombination also tends to be higher toward the ends of chromosomes (*88*), where we found higher GC-rich high-variance exons. An alternative possibility is that the associations of ecology and/or life history are related to convergent exon-coding mutations for those traits in avian genomes (*89*, *90*).

With a well-resolved tree, it becomes possible to more confidently infer evolution of convergent traits. Our tree lends support for either three independent gains of vocal learning (*38*, *91*) or two gains (hummingbirds and the common ancestor of parrots and oscine songbirds) followed by two losses (in New Zealand wrens and suboscines) (*29*, *39*). However, a single origin for parrots and oscines followed by two losses (three events) is not much less parsimonious than independent origins in parrots and oscines (two events). In addition, the suboscine *Procnius* bellbirds have recently been shown to be vocal learners (*92*, *93*), suggesting that there could have been a fourth gain or a regain after a loss of vocal learning in other suboscines. The non-monophyly of the birds of prey at the deepest branches of the Australaves and Afroaves radiations suggests that the common ancestor of core landbirds may have been an apex predator, followed by two losses of the raptorial trait. Seriema at the deepest branch of Australaves could be considered to belong to a raptorial taxon because they kill vertebrate prey (*94*) and are the sole living relatives of the extinct giant "terror birds," apex predators during the Paleogene (*95*, *96*). The deepest branches after Accipitriformes and owl among the Afroaves, the mousebirds and cuckoo-roller, have Eocene relatives with raptor-like feet (*97*), and the cuckoo-roller specializes on chameleon prey (*98*). This suggests that losses of the predatory phenotype were gradual across successive divergences of each of the two core landbird radiations. More

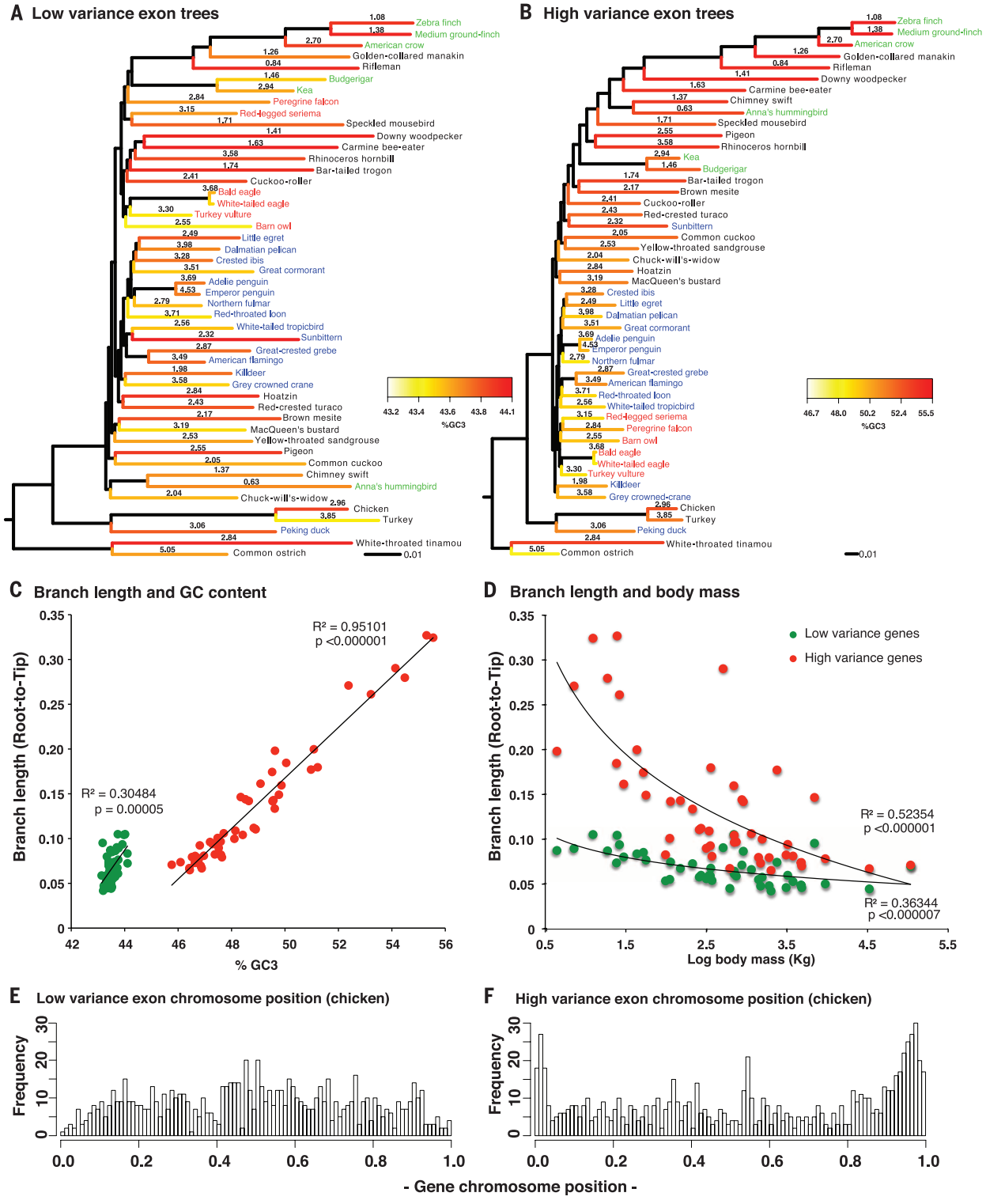


Fig. 6. Life history incongruence in protein-coding trees. (**A**) Species tree inferred from low–base composition variance exons (n = 830 genes) graphed with branch length, third codon position GC (GC3) content (heatmap), and log of body mass (numbers on branches). (**B**) Species tree inferred from high–base composition variance exons (n = 830 genes), graphed similarly as in (A). The %GC3 scale is higher and ~10 times wider for the high-variance genes, and the branch lengths are ~3 times longer [black scales at the bottom of (A) and (B)]. Color coding of species' names is as in Fig. 1. Cladograms of trees in (A) and (B) are in figs. S16, A and B. (**C** and **D**) Correlations of branch length with GC content (C) and body mass (D) of the low-variance and high-variance exons. Correlations were still significant after independent contrast analyses for phylogenetic relationships (SM11). (**E** and **F**) Relative chromosome positions of the low-variance (E) and high-variance (F) exons normalized between 0 and 1 for all chicken chromosomes and separated into 100 bins (bars). The height of each bar represents the number of genes in that specific relative location. The two distributions in (E) and (F) are significantly different ($P < 2.2 \times 10^{-16}$, Wilcoxon rank sum test on grouped values). For further details, see SM11.

broadly, the Columbea and Passerea clades appear to have many ecologically driven convergent traits that have led previous studies to group them into supposed monophyletic taxa (*8*, *17*, *25*). These convergences include the footpropelled diving trait of grebes in Columbea with loons and cormorants (*15*) in Passerea, the wading-feeding trait of flamingos in Columbea with ibises and egrets (*24*, *99*) in Passerea, and pigeons and sandgrouse in Columbea with shorebirds (killdeer) in Passerea (*24*). These long-known trait and morphological alliances suggest that some of the traditional nongenomic trait classifications are based on polyphyletic assemblages.

In conclusion, our genome-scale analysis supports the hypothesis of a rapid radiation of diverse species occurring within a relatively short period of time (36 lineages within 10 to 15 million years; Fig. 1) during the K-Pg transition, with many interordinal divergences in the 1- to 3-million-year range. This rate of divergence is consistent with modern speciation rates, but it is notable that so many lineages from a single stem lineage survived extinction. Subsequent ecological diversification of surviving lineages is consistent with a proliferation of the earliest fossil stem representatives of most modern orders by the latest Paleocene to Eocene. Our finding is broadly consistent with recent estimates for placental mammals [(*100*), but see SM12 (*101*)] and thus supports the hypothesis that the K-Pg transition was associated with a rapid species radiation caused by a release of ecological niches following the environmental destruction and species extinctions linked to an asteroid impact (*2*, *4*, *5*, *102*).

REFERENCES AND NOTES

1. C. Venditti, A. Meade, M. Pagel, *Nature* **463**, 349–352 (2010).
2. J. B. Yoder *et al*., *J. Evol. Biol.* **23**, 1581–1596 (2010).
3. A. Feduccia, *Science* **267**, 637–638 (1995).
4. P. Schulte *et al*., *Science* **327**, 1214–1218 (2010).
5. N. R. Longrich, T. Tokaryk, D. J. Field, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 15253–15257 (2011).
6. D. T. Ksepka, C. A. Boyd, *Paleobiology* **38**, 112–125 (2012).
7. K. M. Cohen, S. Finney, P. L. Gibbard, International Chronostratigraphic Chart v2013/01, in *International Commission on Stratigraphy* (2013); www.stratigraphy.org/ICSchart/ChronostratChart2013-01.jpg.
8. P. G. Ericson *et al*., *Biol. Lett.* **2**, 543–547 (2006).
9. R. W. Meredith *et al*., *Science* **334**, 521–524 (2011).
10. W. Jetz, G. H. Thomas, J. B. Joy, K. Hartmann, A. O. Mooers, *Nature* **491**, 444–448 (2012).
11. O. Haddrath, A. J. Baker, *Proc. Biol. Sci.* **279**, 4617–4625 (2012).
12. M. S. Lee, A. Cau, D. Naish, G. J. Dyke, *Syst. Biol.* **63**, 442–449 (2014).
13. J. W. Brown, J. S. Rest, J. García-Moreno, M. D. Sorenson, D. P. Mindell, *BMC Biol.* **6**, 6 (2008).
14. M. A. Pacheco *et al*., *Mol. Biol. Evol.* **28**, 1927–1942 (2011).
15. B. C. Livezey, R. L. Zusi, *Zool. J. Linn. Soc.* **149**, 1–95 (2007).
16. G. Mayr, *J. Zool. Syst. Evol. Res.* **49**, 58–76 (2011).
17. S. J. Hackett *et al*., *Science* **320**, 1763–1768 (2008).
18. R. C. Pratt *et al*., *Mol. Biol. Evol.* **26**, 313–326 (2009).
19. B. Nabholz, A. Künstner, R. Wang, E. D. Jarvis, H. Ellegren, *Mol. Biol. Evol.* **28**, 2197–2210 (2011).
20. O. Jeffroy, H. Brinkmann, F. Delsuc, H. Philippe, *Trends Genet.* **22**, 225–231 (2006).
21. S. Song, L. Liu, S. V. Edwards, S. Wu, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 14942–14947 (2012).
22. J. Alföldi, K. Lindblad-Toh, *Genome Res.* **23**, 1063–1068 (2013).
23. J. L. Peters, *Check-List of Birds of the World*, E. Mayr, R. A. Paynter, M. A. Traylor, J. C. Greenway, Eds. (Harvard Univ. Press, Cambridge, MA, 1937–1987).
24. C. G. Sibley, J. E. Ahlquist, *Phylogeny and Classification of Birds: A Study in Molecular Evolution* (Yale Univ. Press, New Haven, CT, 1990).
25. M. G. Fain, P. Houde, *Evolution* **58**, 2558–2573 (2004).
26. N. Wang, E. L. Braun, R. T. Kimball, *Mol. Biol. Evol.* **29**, 737–750 (2012).
27. R. T. Kimball, N. Wang, V. Heimer-McGinn, C. Ferguson, E. L. Braun, *Mol. Phylogenet. Evol.* **69**, 1021–1032 (2013).
28. J. E. McCormack *et al*., *PLOS ONE* **8**, e54848 (2013).
29. A. Suh *et al*., *Nat. Commun.* **2**, 443 (2011).
30. M. T. Mahmood, P. A. McLenachan, G. C. Gibb, D. Penny, *Genome Biol. Evol.* **6**, 326–332 (2014).
31. A. Matzke *et al*., *Mol. Biol. Evol.* **29**, 1497–1501 (2012).
32. J. L. Chojnowski, R. T. Kimball, E. L. Braun, *Gene* **410**, 89–96 (2008).
33. S. Patel, R. T. Kimball, E. L. Braun, *J. Phylogenetics Evol. Biol.* **1**, 9–10 (2013).
34. Y. I. Wolf, I. B. Rogozin, N. V. Grishin, E. V. Koonin, *Trends Genet.* **18**, 472–479 (2002).
35. A. Rokas, B. L. Williams, N. King, S. B. Carroll, *Nature* **425**, 798–804 (2003).
36. J. Cracraft, in *The Howard and Moore Complete Checklist of the Birds of the World*, E. C. Dickinson, J. V. Remsen, Eds. (Aves Press, Eastbourne, UK, 2013), pp. xxi–xliii.
37. E. C. Dickinson, J. V. Remsen, Eds., *The Howard and Moore Complete Checklist of Birds of the World* (Aves Press, Eastbourne, UK, 2013).
38. E. D. Jarvis, *Ann. N.Y. Acad. Sci.* **1016**, 749–777 (2004).
39. D. F. Clayton, C. N. Balakrishnan, S. E. London, *Curr. Biol.* **19**, R865–R873 (2009).
40. L. W. Hillier *et al*., *Nature* **432**, 695–716 (2004).
41. R. A. Dalloul *et al*., *PLOS Biol.* **8**, e1000475 (2010).
42. W. C. Warren *et al*., *Nature* **464**, 757–762 (2010).
43. J. Alföldi *et al*., *Nature* **477**, 587–591 (2011).
44. G. Zhang *et al*., *Science* **346**, 1311–1320 (2014).
45. M. D. Shapiro *et al*., *Science* **339**, 1063–1067 (2013).
46. X. Zhan *et al*., *Nat. Genet.* **45**, 563–566 (2013).
47. Y. Huang *et al*., *Nat. Genet.* **45**, 776–783 (2013).
48. Z. Wang *et al*., *Nat. Genet.* **45**, 701–706 (2013).
49. G. Ganapathy *et al*., *Gigascience* **3**, 11 (2014).
50. S. Li *et al*., *Genome Biol.* 10.1186/s13059-014-0557-1 (2014).
51. C. Li *et al*., *GigaScience* **3**, 27 (2014).
52. R. E. Green *et al*., *Science* **346**, 1254449 (2014).
53. K. Liu, S. Raghavan, S. Nelesen, C. R. Linder, T. Warnow, *Science* **324**, 1561–1564 (2009).
54. K. Liu *et al*., *Syst. Biol.* **61**, 90–106 (2012).
55. A. Stamatakis *et al*., *Bioinformatics* **28**, 2064–2066 (2012).
56. J. Zhang, A. Stamatakis, "The multi-processor scheduling problem in phylogenetics," *IEEE 26th International Parallel and Distributed Processing Symposium*, Shanghai, 21 to 25 May 2012, pp. 691–698; 10.1109/IPDPSW.2012.86.
57. A. Stamatakis, A. J. Aberer, "Novel parallelization schemes for large-scale likelihood-based phylogenetic inference," *IEEE 27th International Symposium on Parallel and Distributed Processing*, Boston, 20 to 24 May 2013, pp. 1195–1204; 10.1109/IPDPS.2013.70.
58. S. Mirarab, M. S. Bayzid, B. Boussau, T. Warnow, *Science* **346**, 1250463 (2014).
59. T. Yuri *et al*., *Biology (Basel)* **2**, 419–444 (2013).
60. P. G. Ericson, *J. Biogeogr.* **39**, 813–824 (2012).
61. J. Kim, *Mol. Phylogenet. Evol.* **17**, 58–75 (2000).
62. W. P. Maddison, *Syst. Biol.* **46**, 523–536 (1997).
63. N. A. Rosenberg, *Mol. Biol. Evol.* **30**, 2709–2713 (2013).
64. J. H. Degnan, N. A. Rosenberg, *Trends Ecol. Evol.* **24**, 332–340 (2009).
65. A. W. Edwards, *Genetics* **183**, 5–12 (2009).
66. M. S. Bayzid, T. Warnow, *Bioinformatics* **29**, 2277–2284 (2013).
67. L. Liu, L. Yu, S. V. Edwards, *BMC Evol. Biol.* **10**, 302 (2010).
68. J. Huerta-Cepas, S. Capella-Gutiérrez, L. P. Pryszcz, M. Marcet-Houben, T. Gabaldón, *Nucleic Acids Res.* **42**, D897–D902 (2014).
69. M. P. Simmons, H. Ochoterena, *Syst. Biol.* **49**, 369–381 (2000).
70. J. Felsenstein, *Inferring Phylogenies* (Sinauer Associates, Sunderland, MA, 2004).
71. J. C. Avise, T. J. Robinson, *Syst. Biol.* **57**, 503–507 (2008).
72. D. A. Ray, J. Xing, A. H. Salem, M. A. Batzer, *Syst. Biol.* **55**, 928–935 (2006).
73. K. L. Han *et al*., *Syst. Biol.* **60**, 375–386 (2011).
74. E. M. Lemmon, A. R. Lemmon, *Annu. Rev. Ecol. Evol.* **44**, 99–121 (2013).
75. C. C. Weber, B. Boussau, J. Romiguier, E. D. Jarvis, H. Ellegren, *Genome Biol.* **15**, 549 (2014).
76. C. C. Weber, B. Nabholz, J. Romiguier, H. Ellegren, *Genome Biol.* **15**, 542 (2014).
77. M. dos Reis, Z. Yang, *Mol. Biol. Evol.* **28**, 2161–2172 (2011).
78. J. V. Smith, E. L. Braun, R. T. Kimball, *Syst. Biol.* **62**, 35–49 (2013).
79. P. Houde, *Nature* **324**, 563–565 (1986).
80. G. Mayr, *Syst. Biodivers.* **11**, 7–13 (2013).
81. J. A. Clarke, C. P. Tambussi, J. I. Noriega, G. M. Erickson, R. A. Ketcham, *Nature* **433**, 305–308 (2005).
82. L. Salichos, A. Rokas, *Nature* **497**, 327–331 (2013).
83. L. Salichos, A. Stamatakis, A. Rokas, *Mol. Biol. Evol.* **31**, 1261–1271 (2014).
84. A. D. Twyford, R. A. Ennos, *Heredity (Edinb.)* **108**, 179–189 (2012).
85. L. Duret, N. Galtier, *Annu. Rev. Genomics Hum. Genet.* **10**, 285–311 (2009).
86. C. F. Mugal, P. F. Arndt, H. Ellegren, *Mol. Biol. Evol.* **30**, 1700–1712 (2013).
87. J. Romiguier, V. Ranwez, E. J. Douzery, N. Galtier, *Genome Res.* **20**, 1001–1009 (2010).
88. M. K. Rudd *et al*., *PLOS Genet.* **3**, e32 (2007).
89. J. Parker *et al*., *Nature* **502**, 228–231 (2013).
90. Proteome-wide analyses among the 48 bird genomes were conducted in (*44*), where it was found that vocal-learning bird species have convergent amino acid changes in a set of genes expressed in the song-learning nuclei.
91. F. Nottebohm, *Am. Nat.* **106**, 116–140 (1972).
92. V. Saranathan, D. Hamilton, G. V. Powell, D. E. Kroodsma, R. O. Prum, *Mol. Ecol.* **16**, 3689–3702 (2007).
93. D. Kroodsma *et al*., *Wilson J. Ornithol.* **125**, 1–14 (2013).
94. K. H. Redford, G. Peters, *J. Field Ornithol.* **57**, 261–269 (1986).
95. H. M. F. Alvarenga, E. Hofling, *Papeis Avulsos Zool.* **43**, 55–91 (2003).
96. G. Mayr, in *Paleogene Fossil Birds* (Springer, Berlin, Heidelberg, 2009), pp. 139–152.
97. P. Houde, S. L. Olson, in *Natural History Museum of Los Angeles County Science Series* (Natural History Museum of Los Angeles County, Los Angeles, 1992), vol. 36, pp. 137–160.
98. A. D. Forbes-Watson, *Ibis* **109**, 425–430 (1967).
99. S. L. Olson, A. Feduccia, *Smithson. Contrib. Zool.* **316**, 1–73 (1980).
100. M. A. O'Leary *et al*., *Science* **339**, 662–667 (2013).
101. M. dos Reis, P. C. Donoghue, Z. Yang, *Biol. Lett.* **10**, 20131003 (2014).
102. M. J. Benton, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **365**, 3667–3679 (2010).

ACKNOWLEDGMENTS

Genome assemblies, annotations, alignments, tree files, and other data sets used or generated in this study are available at GigaScience, the National Center for Biotechnology Information (NCBI), ENSEMBL, CoGe, UCSC, and other sources listed in SM13 (table S17). We thank S. Edmunds at GigaScience, K. Pruit at NCBI, and P. Flicek at ENSEMBL for making this possible. The majority of genome sequencing and annotation was supported by internal funding from BGI. Additional major support is from the coordinators of the project: E.D.J. from the HHMI and NIH Directors Pioneer Award DP1OD000448; S.M. from an HHMI International Student Fellowship; G.Z. from Marie Curie International Incoming Fellowship grant (300837); T.W. from NSF DEB 0733029, NSF DBI 1062335, and NSF IR/D program; and M.T.P.G. from a Danish National Research Foundation grant (DNRF94) and a Lundbeck Foundation grant (R52-A5062). J. Fjeldså generated the bird drawings used in the figures. O.A.R. acknowledges a uniform biological material transfer agreement between San Diego Zoo Global and BGI used for some tissue samples. R.E.G. declares that he is President of Dovetail Genomics, with no conflicts of interest. Additional acknowledgements are listed in the supplementary materials. We thank the following for allowing us to conduct the computationally intensive analyses for this study: Heidelberg Institute for Theoretical Studies; San Diego Supercomputer Center, with support by an NSF grant; SuperMUC Petascale System at the Leibniz Supercomputing Center; Technical University of Denmark; Texas Advanced Computing Center; Georgia Advanced Computing Resource Center, a partnership between the University of Georgia's Office of the Vice President for Research and Office of the Vice President for Information Technology; Amazon Web Services; BGI; the Nautilus supercomputer at the National Institute for Computational Sciences of the University of Tennessee and Smithsonian; and Duke University Institute for Genome Sciences and Policy.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/346/6215/1320/suppl/DC1
Supplementary Text SM1 to SM13
Figs. S1 to S29
Tables S1 to S17
References (*103*–*277*)
Appendices

17 March 2014; accepted 15 October 2014
10.1126/science.1253451

RESEARCH ARTICLE SUMMARY

AVIAN GENOMICS

Complex evolutionary trajectories of sex chromosomes across bird taxa

Qi Zhou,* Jilin Zhang, Doris Bachtrog, Na An, Quanfei Huang, Erich D. Jarvis, M. Thomas P. Gilbert, Guojie Zhang*

INTRODUCTION: Sex chromosomes originate from ordinary autosomes. Ancient sex-specific W or Y chromosomes, like that of female chicken or those of male mammals, usually have lost most functional genes, owing to a loss of recombination with their former homologs, the Z or X chromosomes. Such a recombination restriction occurred in a stepwise manner along most of the sex chromosomes in parallel in birds and mammals (creating so-called "evolutionary strata"), apart from small pseudoautosomal regions (PARs) that maintain recombination. Sex chromosomes of some basal birds like ostrich and emu have exceptionally large recombining PARs, but little is known about the genomic composition of most bird species' sex chromosomes.

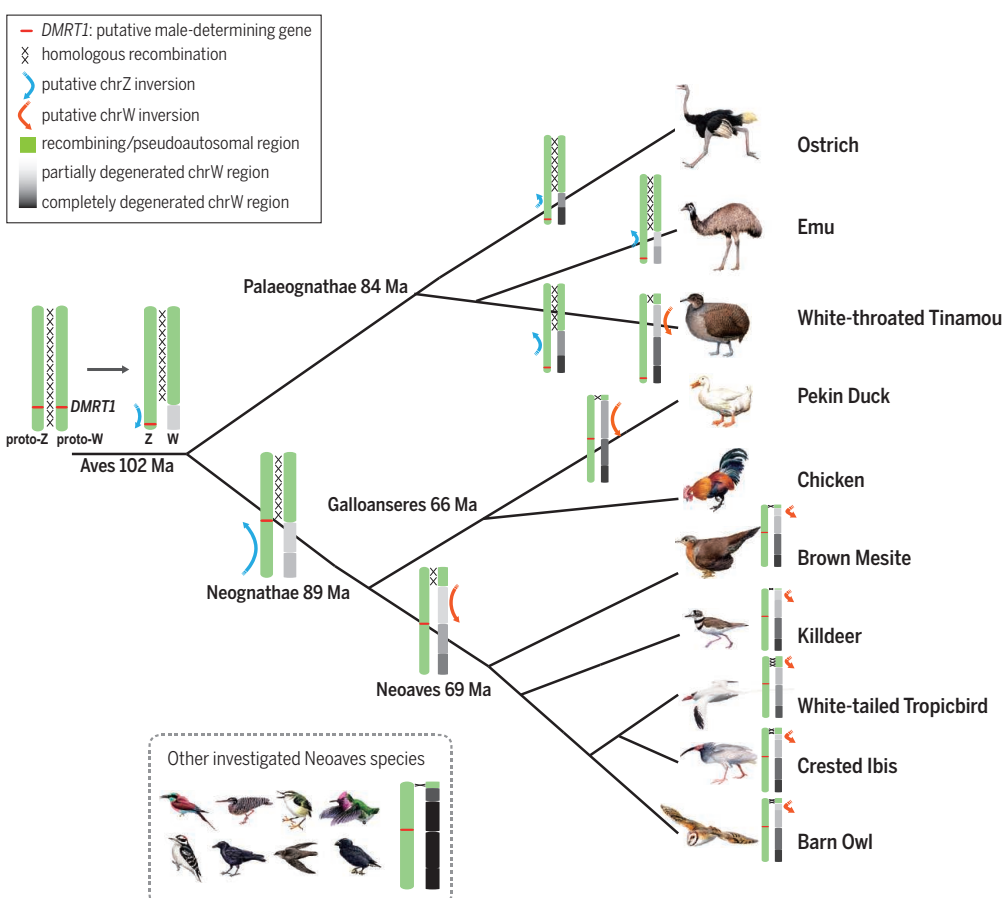
RATIONALE: Here we use the newly available genomes of 17 species spanning the entire avian phylogeny to decipher the genomic architecture and evolutionary history of bird sex chromosomes. We demarcate the PAR and the nonrecombining differentiated region between Z/W of each species by their different read depths relative to autosomes. We further assemble numerous W-linked genomic regions, whose abundance and sequence divergence level with Z chromosome reflect their ages of recombination loss.

RESULTS: Surprisingly, we find that more than half of the studied species have a W chromosome that is not completely degenerated. Besides ostrich and emu, some Neognathae species like tropicbird and killdeer also have long PARs. The nonrecombining regions between Z/W of many species exhibit a complex pattern of "evolutionary strata," resulting from the suppression of recombination in a stepwise and independent manner among some lineages. We conclude that the first evolutionary stratum that contains the putative male-determining gene *DMRT1* formed through a Z-linked chromosome inversion in an ancestor of all birds. This was followed by one stratum formed in the ancestor of Neognathae, one stratum in the ancestor of Neoaves, and independent emergence of more recent strata in most species. Many W-linked genes have disrupted protein function or reduced gene expression, and the rate of functional decay significantly slows down in older strata.

CONCLUSION: Our study uncovered an unexpected complexity of avian sex chromosomes, due to the lineage-specific recombination suppressions and different tempo of W degeneration. In contrast to mammals, some birds never experienced global recombination arrest, or differentiate at a very low rate between Z/W even after the recombination loss. This may relate to different intensities of sexual selection across bird species and explain their lack of a general chromosome-wide dosage compensation mechanism. ■

ON OUR WEB SITE

Read the full article at http://dx.doi.org/10.1126/science.1246338



Evolutionary strata of avian sex chromosomes. Bird sex chromosomes suppressed homologous recombination through shared or lineage-specific chromosomal inversions. Each defines an "evolutionary stratum" that exhibits a distinctive level of Z/W sequence divergence from neighboring strata. We demarcated and dated each stratum and show that bird sex chromosomes harbor great diversity regarding the lengths of recombining regions (green bars with crosses), or the degree of Z/W differentiation (scaled from gray to black).

The list of author affiliations is available in the full article online.
*Corresponding author. E-mail: zhouqi@berkeley.edu (Q.Z.); zhanggj@genomics.org.cn (G.Z.)
Cite this article as Q. Zhou *et al.*, *Science* **346**, 1246338 (2014). DOI: 10.1126/science.1246338

Downloaded from www.sciencemag.org on December 14, 2014

Published by AAAS

RESEARCH ARTICLE

Complex evolutionary trajectories of sex chromosomes across bird taxa

Qi Zhou,[1*†] Jilin Zhang,[2*] Doris Bachtrog,[1*] Na An,[2] Quanfei Huang,[2] Erich D. Jarvis,[3] M. Thomas P. Gilbert,[4,5] Guojie Zhang[2,6†]

Sex-specific chromosomes, like the W of most female birds and the Y of male mammals, usually have lost most genes owing to a lack of recombination. We analyze newly available genomes of 17 bird species representing the avian phylogenetic range, and find that more than half of them do not have as fully degenerated W chromosomes as that of chicken. We show that avian sex chromosomes harbor tremendous diversity among species in their composition of pseudoautosomal regions and degree of Z/W differentiation. Punctuated events of shared or lineage-specific recombination suppression have produced a gradient of "evolutionary strata" along the Z chromosome, which initiates from the putative avian sex-determining gene *DMRT1* and ends at the pseudoautosomal region. W-linked genes are subject to ongoing functional decay after recombination was suppressed, and the tempo of degeneration slows down in older strata. Overall, we unveil a complex history of avian sex chromosome evolution.

In many species with separate sexes, sex is determined by a pair of heteromorphic sex chromosomes that differ in their size, morphology, and gene content. Mammals have male heterogametic sex chromosomes (XX in female, XY in male), whereas birds have female heterogametic sex chromosomes (ZW in female and ZZ in male) (*1*). Both sex systems have originated independently from different ancestral autosomes after the two lineages diverged more than 300 million years ago (Ma) (*1*–*4*), yet they share many common features with respect to their genomic composition and evolutionary history. Most regions of mammalian Y or avian W chromosomes (chrW) lack recombination except for the pseudoautosomal regions (PARs) that are shared between X and Y, or Z and W, and which encompass only 2% of the human X (*5*). The nonrecombining region of the Y/W usually has lost almost all functional genes present on the ancestral chromosome except for a few loci, often with sex-specific functions (*6*), and Y/W chromosomes are often largely composed of repetitive and epigenetically silenced (heterochromatic) DNA (*7*). Recombination suppression is thought to follow the origination of a sex-determining locus and may initially encompass only a small chromosomal segment surrounding that locus, but then progressively spreads along Y or W chromosomes, possibly owing to the accumulation of sexually antagonistic alleles at PARs that benefit one sex but harm the other (*8*). Chicken and human sex chromosomes have in parallel evolved recombination suppression in a stepwise manner, manifested as "evolutionary strata" with a gradient of ages. Each stratum is characterized by a distinct range of levels of sequence homology between the remaining XY or ZW gene pairs, reflecting the punctuated evolutionary time points when recombination between the sex chromosomes ceased (*9*, *10*). To date, genomic sequences of 10 mammalian species' Y chromosomes have been determined (*11*, *12*), and the human Y was found to have five strata (*5*) probably resulting from Y-linked inversions (*13*), of which three are shared among all Eutherian mammals (*4*, *14*). No complete avian chrW has been sequenced, but partial annotation of W-linked genes in chicken (*Gallus gallus*) (*15*) implied that there are at least three strata (*10*, *16*), and some may have formed independently in different avian lineages (*17*–*20*).

A recent investigation of the karyotype of 200 avian species revealed great variation in the size and morphology of chrW, sometimes even between closely related species (*21*), suggesting a dynamic evolutionary history of bird sex chromosomes. Living birds consist of two sister lineages, the Palaeognathae (flightless ratites and Tinamous) and the Neognathae, which themselves consist of two major lineages, the Galloanserae (land and water fowls) and the Neoaves (all other neognaths). Neoaves are derived from an ancient species radiation and comprise more than 95% of all avian species (*22*, *23*). Karyotyping suggests that most ratite paleognaths, like ostrich (*Struthio camelus*) and emu (*Dromaius novaehollandiae*), have homomorphic sex chromosomes with extensive recombining and euchromatic PARs that resemble the ancestral state (*24*–*26*). In contrast, their sister group Tinamous shows interspecific variation in sizes and locations of PARs (*27*). Whether all Neognathae species have an almost completely degenerated chrW similar to that of chicken (*28*) is unknown. Here, we analyzed female genomes of 17 bird species newly generated by the Avian Genomics Consortium that span the avian taxonomic diversity (*23*, *29*), in order to unveil the broad evolutionary history of bird ZW chromosomes.

Results

Avian sex chromosomes have different lengths of PAR

We collected genome sequences of three Palaeognathae (ostrich, emu, white-throated tinamou), one Galloanserae (Pekin duck) (*30*), and 13 Neoaves species, including a representative of the sister lineage to all other Neoaves (brown mesite), the Caprimulgiformes (Anna's hummingbird and chimney swift), core waterbirds (crested ibis to white-tailed tropicbird), and core landbirds (rifleman to woodpecker) (*29*) (Fig. 1 and table S1). For three species—chicken, emu, and crested ibis—we analyzed sequence data of both sexes (*31*–*33*). We additionally improved the ostrich draft genome to a chromosome assembly with optical mapping (*34*) (N50: 17.7 Mb, table S2). Comparative cytogenetic mapping and genome sequence analysis show that there are very few interchromosomal rearrangements among bird species (*35*–*38*); thus, we anchored genome scaffold sequences of the remaining species using the ostrich Z chromosome (chrZ) for all Palaeognathae and chicken Z for all Neognathae (*3*).

To distinguish between the recombining PAR (that has nearly identical sequences between chrZ and chrW) and the region lacking recombination (that has become differentiated between chrZ and chrW) along the sex chromosome, we developed an approach that involves examination of read sequencing depth along chrZ (*34*). Although most sequences in the female-specific W-linked region may have lost almost all homology with chrZ, some ZW homolog pairs (often called gametologs) may remain. However, W-linked regions are probably too divergent to allow short sequencing reads derived from them [100 base pairs (bp) long in this study] to align with their Z homologs, especially if they are from more ancient strata. Consequently, PARs will show a level of read depth similar to that of autosomes, whereas differentiated regions will have only half the mapped read coverage depth in females (Fig. 1).

Based on this principle, we find evidence of extreme differences in the length of the PAR versus differentiated region on sex chromosomes across species (Fig. 1 and Table 1), in contrast to a lack of such variation in coverage from autosomes (fig. S1). The putative avian male sex-determining gene *DMRT1* (*39*) always resides in the differentiated region in all species, consistent with it being the initial seed for progressive recombination suppression (Fig. 1). However, the size of the PAR varies substantially among species (Table 1). In Palaeognathae, at least two-thirds of chrZ of ostrich and emu are pseudoautosomal and still

[1]Department of Integrative Biology, University of California, Berkeley, CA94720, USA. [2]China National Genebank, BGI-Shenzhen, Shenzhen, 518083. China. [3]Department of Neurobiology, Howard Hughes Medical Institute, Duke University Medical Center, Durham, NC 27710, USA. [4]Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5-7, 1350 Copenhagen, Denmark. [5]Trace and Environmental DNA laboratory, Department of Environment and Agriculture, Curtin University, Perth, Western Australia 6102, Australia. [6]Centre for Social Evolution, Department of Biology, Universitetsparken 15, University of Copenhagen, DK-2100 Copenhagen, Denmark.
*These authors contributed equally to this work. †Corresponding author. E-mail: zhouqi@berkeley.edu (Q.Z.); zhanggj@genomics.org.cn (G.Z.)

recombine with chrW, in contrast to only 1% of chrZ of white-throated tinamou (*Tinamus guttatus*) being PAR (Fig. 1 and Table 1). We confirm that we have identified the true PAR in ostrich, as we find that eight genes that have been cytologically mapped to both chrZ and chrW (*26*) are appropriately distributed along our assembled ostrich chrZ (Fig. 1). Whereas most Neognathae have restricted recombination between their sex chromosomes into only a very short terminal PAR (Fig. 1 and fig. S2), unexpectedly two Neoavian species, killdeer (*Charadrius vociferus*) and white-tailed tropicbird (*Phaethon lepturus*), retain PARs of considerable length, encompassing between 5 and 9% of their chrZ (Table 1). Thus, unlike chicken (*40*), some Neognathae chrW can pair and recombine with chrZ through a sizable PAR. This also indicates that recombination suppression in the genomic segment adjacent to the PAR has evolved independently in at least some bird lineages (see below).

To further quantify Z/W divergence levels within the differentiated region, we searched for assembled W-linked sequences longer than 1 kb, which allow better alignments with chrZ than the raw reads. W gametologs may be embedded within repetitive DNA that cannot be readily assembled into large scaffolds, but unique segments of a species' chrW are expected to assemble into fragmented scaffold sequences separate from those of chrZ. If these W-segments are only differentiated to a moderate degree, they may have retained sufficient sequence homology to map to the differentiated region along chrZ. Using this approach, we identified candidate W-linked scaffolds up to 25% of the homologous chrZ length with low background mapping to autosomes, indicating effective discrimination from random fragments (*34*) (fig. S1). We identify abundant W-linked sequences from all three Palaeognathae species, one of the two Galloanserae species and in five of the Neoaves (Table 1 and table S3), and many fewer candidate W-derived fragments homologous to chrZs in chicken and the remaining Neoaves species (Fig. 1 and fig. S2).

Comparison of read depths between the two sexes in chicken, crested ibis, and emu verified that the candidate W-linked scaffolds are specific to females (figs. S3 and S4). Additionally, most of the candidate W-linked scaffolds of chicken exclusively align to W-linked or unmapped (chrUn) fragments of the reference chicken genome derived from Sanger sequencing (fig. S5). These results together confirmed that we have identified truly W-linked scaffolds. We further annotated genes in these W-derived scaffolds, ranging from 24 in the Pekin duck to 55 in ostrich and 198 in the white-throated tinamou (Table 1 and table S4), compared to 26 and 14 W-linked genes previously annotated through transcriptome sequencing in chicken and ostrich, respectively (figs. S6 and 7 and table S5) (*15*, *41*). Some W-linked genes may be missing or fragmented (table S5), if they are highly enriched for transposable elements or too similar to their Z-linked homologs in sequence. Nevertheless, we found the coding sequence lengths of the Z/W gene pairs identified in this work to generally correlate well with those of their chicken Z-linked orthologs (fig. S8). This indicates that, although our current chrW genome assemblies are interrupted by highly repetitive noncoding regions, they contain substantial amounts of coding regions of genes. Overall, we find that avian chrWs show great variation in their degree of degeneration, and more than one-third of the sampled Neognathae species harbor chrWs that are not as fully degenerated as that of chicken (fig. S2).

Evolutionary strata of Palaeognathae species

The identified W-linked fragments are stratified by differences in their occurrences and alignment identities along chrZ (figs. S9 and S10), exhibiting patterns of evolutionary strata. This provides evidence that, similar to those of mammals (*4*), avian sex chromosomes suppressed recombination

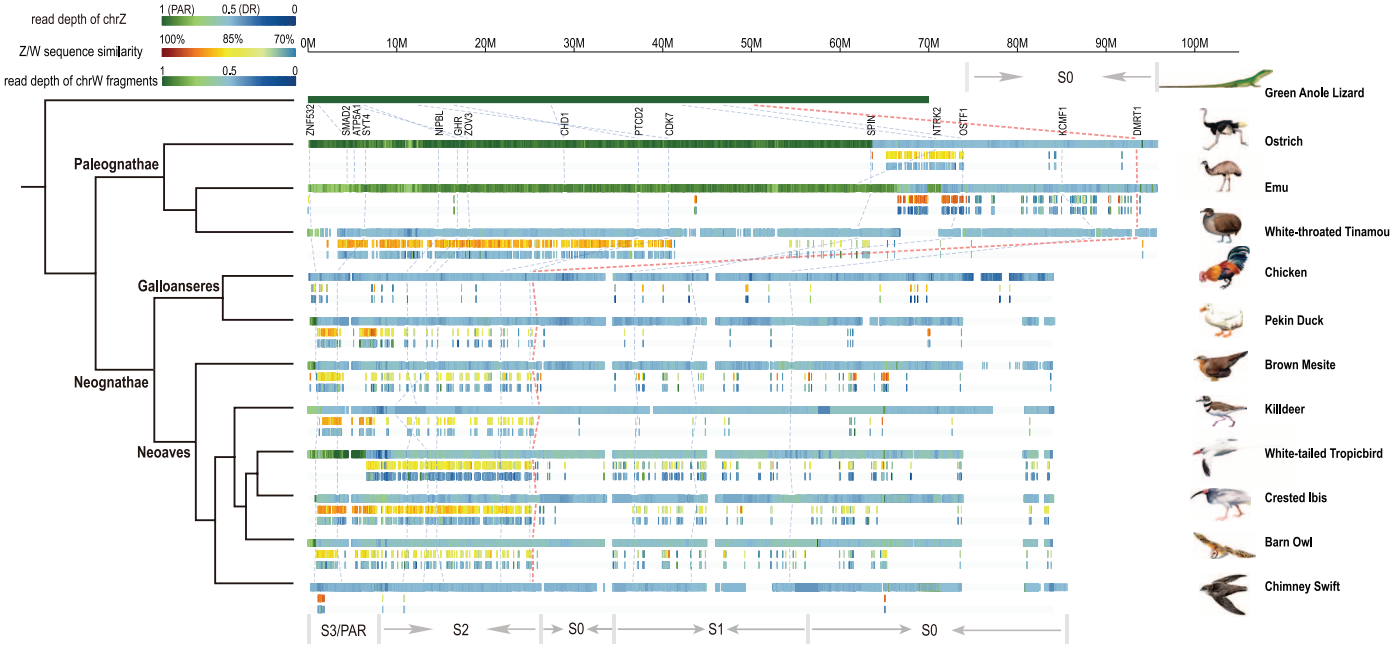


Fig. 1. Evolutionary strata of avian sex chromosomes. We present evolutionary strata patterns of investigated bird species with partially degenerated chrW, except for chicken and chimney swift as comparison, together with their revised phylogeny (*23*). A full pattern for all the investigated species is presented in fig. S2 and summarized in fig. S23. chrZs of tinamou and emu are ordered and scaled to the same length as that of ostrich, and all others to that of chicken. Each avian species includes three color-coded tracks from top to bottom: (i) female read depth of chrZ, where PAR is shown in green, differentiated region without recombination between Z/W in blue, and assembly gaps as blanks; (ii) pairwise alignment identities between Z/W plotted from 100% identical (red) to 70% identical (blue); and (iii) female read depth of W-linked sequence fragments, color-coded the same as (i). We labeled regions of each stratum at the bottom: "S0" corresponds to the most ancient evolutionary stratum 0 that is shared by all the species, while the younger strata are abbreviated as S1, S2, etc. Note that different species may not share the same S1, S2, and S3; and part of S0 (25 to 35 Mb in the figure) has been reshuffled because of chromosome inversions that occurred in the ancestor of Neognathae (Fig. 3 and fig. S15). The blank region around 75 to 85 Mb corresponds to a chicken-specific amplification of testis gene clusters (*3*) (table S7). We also labeled genes previously used for cytological mapping (*26*) and the reported S0 gene *KCMF1* (*4*) (traced through the phylogeny with blue dotted lines), as well as the putative male sex-determining gene *DMRT1* (traced with the dotted red line) (*39*).

through a series of punctuated events (*10*, *16*, *17*). The strata present a gradient of ages, reflected by higher levels of sequence homology between chrW and chrZ sequences as one moves toward the PAR (Fig. 1 and figs. S9 to S11). The putative male sex-determining gene of birds, *DMRT1* (*39*), is located within the first stratum (we denote this region as stratum S0) that evolved in the ancestor of all extant birds, around 100 (98 to 108) Ma (*23*). This is supported by a lack of abundant assembled W-linked fragments due to the ancient origin of S0, and also *DMRT1*'s close linkage in our new ostrich genome assembly to *KCMF1* (Fig. 1), which was also characterized as an S0 gene by an independent study through transcriptome sequencing (*4*). The boundaries of S0 can still be clearly traced by the nearby stratum in ostrich and emu (Fig. 1), both of which have maintained the most ancestral state of sex chromosome differentiation among the studied birds and share 90% of their Z-linked gene content of S0 (Fig. 2A).

Maximum-likelihood trees constructed with multiple S0 genes (including *KCMF1*) show that Z-linked gametologs of Palaeognathae and Neognathae cluster with each other by species rather than with their W-linked gametologs (fig. S12), confirming that recombination between chrZ and chrW at S0 was abolished in the common ancestor of all birds (*17*) (Fig. 2A and fig. S12). Consistent with their postulated ancient origin, almost no W-linked genes were identified within the orthologous region of S0 in all the studied species (Table 1), the exception being emu, whose W-linked S0 region has not completely degenerated (*23*, *31*). All residual W-linked fragments within emu S0 show similar levels of divergence from their homologous chrZ sequences (fig. S9), suggesting that they stopped recombining at the same time.

Adjacent to the S0 stratum, we identified one additional stratum (S1) in ostrich and emu, and at least two strata (S1 and S2) in the white-throated tinamou. In contrast to S0, S1 spans different regions in each of the three paleognaths: Each species shares less than 50% of its S1 gene content with the other two (Fig. 2B), and S2 is specific to tinamou, whose orthologous region in the ostrich and emu is PAR (Figs. 1 and 2C). This indicates that these younger strata formed independently in these three anciently diverged species. A small segment of the ostrich PAR is embedded within the inferred emu S1, which suggests a chromosomal rearrangement between these two birds (Fig. 1). Molecular clock dating with Z/W divergence levels at silent sites suggests that stratum S1 of ostrich formed about 75 (34 to 193) Ma and that of emu formed about 23 (11 to 59) Ma, and their sex chromosomes have maintained their current PARs ever since (table S6). White-throated tinamou has a much smaller PAR due to two separate strata that formed about 50 (21 to 95) Ma and 29 (7 to 82) Ma (table S6), after its divergence with ostrich about 58 to 95 Ma (*23*). Thus, common and lineage-specific events characterize sex chromosome evolution in Palaeognathae birds.

Evolutionary strata of Neognathae species

In comparison, Neognathae species exhibit either three or four strata along their sex chromosomes (Fig. 1). We infer that the oldest Neognathae-specific stratum S1 formed in an ancestor of all Neognathae adjacent to the S0 region described above (Fig. 1). We also recovered W-linked sequence fragments within this stratum by reordering them against the ostrich chrZ as a proxy for the proto-Z (fig. S13). Consistent with its ancestral emergence and previous results in chicken (*16*), Neognathae S1 harbors almost no identifiable W-linked fragments or genes in Galloanserae, and we find only few W-linked fragments with low sequence identities versus chrZ in Neoaves (figs. S9 and S10). The low density of W fragments within Neognathae S1 hampers us from determining the exact genomic boundaries of this stratum, but phylogenetic trees constructed from Z/W gametologs of S1 confirm that this stratum is shared among all Neognathae (Fig. 2D and fig. S14). Three Z/W gene pairs, *SPIN*, *HINT*, and *CHD1*, which were previously inferred to have stopped recombining before the Neoaves–Galloanserae split (*17*), are consistently included within the S1 region of Neognathae, despite the shuffling of gene order due to chromosomal rearrangements (fig. S15). By scaling the Z/W divergence level in noncoding regions to that of S2 (see below), we estimate that Neognathae S1 was formed about 71 to 119 Ma, which is consistent with its formation before the inferred split between Neoaves–Galloanserae about 83 to 94 Ma (*23*) (table S6).

Adjacent to S1, we identify a younger Neognathae stratum S2 that spans a similar region in all Neognathae (Fig. 1), but Z/W gametologs within this region cluster by lineage rather than by chromosomes (Fig. 2E and fig. S16), suggesting the Galliformes (chicken), Anseriformes (Pekin duck),

Table 1. Age and gene content of avian evolutionary strata.

Stratum	Species	Ostrich	Emu	White-throated tinamou	Pekin duck	Brown mesite	Killdeer	White-tailed tropicbird	Crested ibis	Barn owl
S0	Age (My)*					102				
	Z-linked genes	247	224	244	154	137	96	157	182	151
	W-linked genes†	0	6/8	0	0	1/1	0/0	2/3	5/4	1/2
	W/Z length (Mb)	0/21.4	1.2/19.8	0/18.3	0.01/32.5	0.2/18.4	0.1/31.9	0.8/20.9	0.8/28.5	0.3/19.4
S1‡	Age (My)	75	23	50			71–119			
	Z-linked genes	108	21	201	333	236	400	246	267	211
	W-linked genes	36/19	0/0	21/9	0/4	7/3	1/1	10/10	8/5	5/4
	W/Z length (Mb)	5.1/10.6	0.3/3.6	1.2/17.7	0.09/16.9	0.08/25.0	0.06/20.3	0.2/26.7	0.5/23.8	0.09/26.8
S2	Age (My)			29	92			69		
	Z-linked genes	PAR		461	237	256	279	224	229	203
	W-linked genes			94/74	14/9	23/8	27/11	37/19	33/32	12/10
	W/Z length (Mb)			16.6/35.1	0.5/17.7	1.6/19.8	2.2/21.2	5.0/19.1	6.3/20.0	1.7/17.8
S3	Age (My)				60	47	45		53	50
	Z-linked genes	PAR		PAR	96	21	24	PAR	25	36
	W-linked genes				13/5	7/3	3/3		2/1	5/2
	W/Z length (Mb)				1.3/8.3	1.5/3.7	2.4/4.7		0.6/3.2	1.1/6.2
PAR	Z-linked genes	718	619	12	18	19	37	55	8	13
	Z lengths (Mb)	63.6	65.5	0.7	1.05	0.7	4.2	6.6	0.4	0.9
Total	**W/Z genes**	55/1073	14/864	198/918	45/848	53/669	46/836	81/682	90/711	41/614
	W/Z lengths	5.1/95.6	1.6/88.9	17.9/71.9	1.9/76.5	3.4/67.7	4.7/82.4	6.0/73.2	8.3/75.9	3.3/71.1

*Ages (My, millions of years) were estimated based on Z/W noncoding sequence divergence level with the assumption of a molecular clock except for stratum S0 and Neoaves S2. Confidence intervals for the ages are shown in table S6. Note that strata with the same number are not necessarily shared among different species. † Number of genes with intact ORFs or disrupted ORFs on the W chromosome. ‡ Boundaries of stratum S0/S1 may not be accurate owing to the low number of identifiable W-linked sequences.

and Neoaves may have each independently formed stratum S2, after the shared origination of S1. On the basis of either the proposed divergence times between these lineages (*23*) or from Z/W sequence divergence level in noncoding regions, we estimate that the duck stratum S2 arose 68 (18 to 127) Ma and that of Neoaves arose 69 (66 to 72) Ma (table S6), whereas the chicken S2 was previously estimated to have arisen 34 to 54 Ma (*16*).

One Neoaves species, the white-tailed tropicbird, seems to have experienced no further recombination suppressions after stratum S2, and its current PAR appears to be comparable to the Neoaves ancestor PAR (Fig. 1). By contrast, orthologs of tropicbird PAR genes are located in differentiated regions of all the other Neoaves, including brown mesite (*Mesitornis unicolor*), suggesting that the youngest stratum S3 evolved independently in different lineages. Indeed, only five genes overlap between the different S3 regions (Fig. 2F), and Z-linked genes from S3 cluster with W-linked gametologs from the same species and not with Z-linked orthologs of other species (Fig. 2F and fig. S17). Two Z/W gene pairs within the S3 region, *ATP5A1* and *UBAP2*, were previously shown to have suppressed recombination independently among different bird species (*17*). Together, these results indicate that recombination in S3 ceased in some lineages after the Neoaves species radiation, and we estimate that S3 emerged at different time points in different lineages (between 45 and 53 Ma; Table 1 and table S6), by scaling its Z/W divergence level at noncoding regions to that of Neoaves S2.

Putative Z- and W-linked inversions causing recombination suppression

Recombination suppression at evolutionary strata may have been mediated by chromosomal inversions, as has been suggested for mammals (*9*, *13*). Genomic comparison of four avian species with de novo assembled genomes against the reptile outgroups green anole lizard (*Anolis carolinensis*) (*42*) and boa snake (*Boa constrictor*) (*43*) (Fig. 3 and fig. S18) allow us to infer the presence of inversions on chrZ and further inform on the

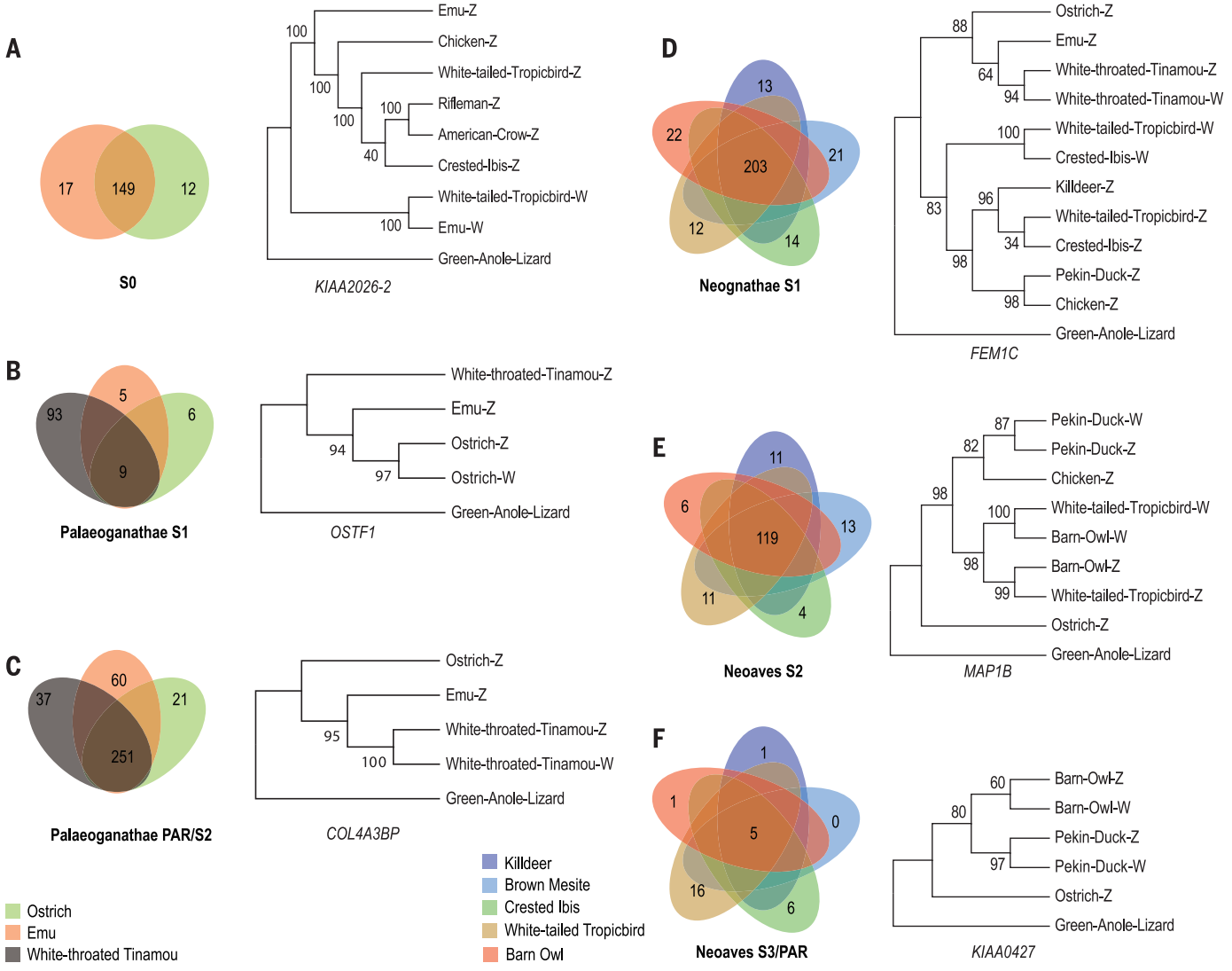


Fig. 2. Emergence of evolutionary strata in different avian lineages. Shown are Venn diagrams of overlapping gene content between orthologous strata of different Palaeognathae (**A** to **C**) and Neognathae (**D** to **F**) species. To the right of each diagram is a maximum-likelihood tree for a specific example gene constructed from Z- and W-linked gametologs of that stratum. Additional trees supporting the inference of a certain stratum's age are shown in the figs. S12, S14, S16, and S17. The tree is constructed with ostrich or green anole lizard as outgroup, with 100 bootstraps. A pattern in which Z gametologs are grouped together by species indicates that Z/W recombination was suppressed before their speciation, whereas clustering of Z and W gametologs of the same species indicates that Z/W recombination was suppressed after the speciation event. "Palaeognathae PAR" shows the number of genes that overlapped between ostrich PAR, emu PAR, and tinamou stratum S2. "Neoaves S3" shows the overlapped gene content between tropicbird PAR versus S3 of four other Neoaves species: killdeer, brown mesite, crested ibis, and barn owl.

formation paths of evolutionary strata in birds. Both cytogenetic mapping (*44*) and our genomic comparison have shown that ostrich has maintained the most ancestral gene synteny relative to other avian species with only a few intrachromosomal rearrangements (Fig. 3). We specifically compared gene synteny spanning the ostrich S0 with the assembled genomes of green anole lizard and boa snake and to cytogenetic data from three other reptiles (rat snake, gecko lizard, Chinese soft-shelled turtle) (*45*), and identified a chromosomal inversion on ostrich chrZ relative to reptiles, which are all collinear within their orthologous regions on the corresponding autosome (fig. S18). This suggests that an inversion on the proto-Z abolished recombination between chrZ and chrW at S0 in an ancestor of all birds. Gene content and synteny within the ostrich S1 region is largely conserved between the green anole lizard and ostrich, suggesting that a putative inversion creating the ostrich-specific S1 stratum must have occurred on chrW (Fig. 3).

Multiple genomic rearrangements within Neognathae S0 and S1 regions among Neognathae species preclude us from inferring the origin of Neognathae S1 (figs. S13 and S15). However, both cytogenetic (*35*) and genomic mapping (Fig. 3) place *DMRT1* close to the telomere of Palaeognathae chrZ but in the middle of chrZ for most other Neognathae species. This suggests that at least one Z-linked inversion, which may have been associated with the formation of Neognathae S1, has relocated *DMRT1* in the ancestor of all Neognathae. This putative inversion has shuffled part of the oldest stratum S0 containing *DMRT1* in between the younger strata Neoaves S2 and S1 in chicken (Fig. 3 and fig. S15), which explains why *DMRT1* was placed in a more recent stratum in a previous study using chicken Z chromosome coordinates as reference (fig. S15) (*16*). Gene synteny is to a large extent conserved across species within both S2 and S3 (Fig. 3 and figs. S18 and S19). This suggests that if inversions are responsible for recombination suppression between chrZ and chrW at S2 and S3, they originated on chrW.

Tempo of W degeneration in birds

Once recombination is suppressed between chrW and chrZ, the functions of W-linked genes are expected to decay (*6*). Y chromosome genome and transcriptome sequencing of 15 mammals has shown a loss of more than 95% of all genes hypothesized to have been present on the ancestral Y of each lineage (*4*, *12*). In contrast, we find extensive heterogeneity across birds in levels of W degeneration, as measured by the proportion of residual W-linked genes or sequences relative to their Z-linked homologs (Table 1). Heterogeneity in W degeneration is found not only between species groups whose strata evolved independently (e.g., Paleognathae versus Neognathae), but also within homologous strata dating back to a common ancestor, and which therefore have carried the same ancestral gene content and underwent degeneration for the same length of time (Fig. 1 and Table 1). This is exemplified by the W-linked region of Neoaves stratum S2 having become almost completely differentiated and lost nearly all genes in 8 out of 13 Neoaves species (fig. S2), whereas tropicbird and crested ibis have retained at least 20% of their ancestral W-linked genes relative to the homologous Z-linked regions (Table 1). Within each stratum, we estimate that chrWs of different species have lost 52 to 95% of their ancestral gene content (Fig. 4, A to G), which may be an overestimate owing to difficulties in assembling W gametologs within highly repetitive regions (figs. S6 and S7). Functional decay is also apparent at assembled W gametologs that are present in the nonrecombining region of chrW (*46*): We find that about 26 to 50% of W-linked genes within each stratum have likely become pseudogenes, owing to the accumulation of premature stop codons and/or frame-shift mutations (Fig. 4, A to G). In addition, W-linked genes in all species consistently show significantly higher rates of protein evolution than their Z-linked gametologs (comparing pairwise rates versus chicken; Wilcoxon test: $P < 0.05$, fig. S20), indicating that they experience less efficient purifying selection on the nonrecombining chrW than the Z gametologs (*46*, *47*). Disabling mutations may have also accumulated in regulatory regions: W-linked gametologs within the differentiated region of ostrich, emu, and duck are expressed at a significantly lower level than both their Z-linked gametologs and lizard orthologs (Wilcoxon test, $P < 0.05$, Fig. 4H and fig. S21) in different tissues, and W gametologs with disrupted open reading frames are down-regulated even more (fig. S22).

Older strata generally have lost more W-linked genes (Fig. 4, A to G), yet we find a significant negative correlation between the rate of gene loss (measured as percent of gene loss per million years) versus the age of stratum in birds ($R^2 = 0.87$, Pearson's test P-value $= 2.87 \times 10^{-12}$, Fig. 4I), similar to those seen in mammalian and chicken strata (*12*, *16*, *49*). Thus, older strata appear to lose genes at a lower rate in both clades,

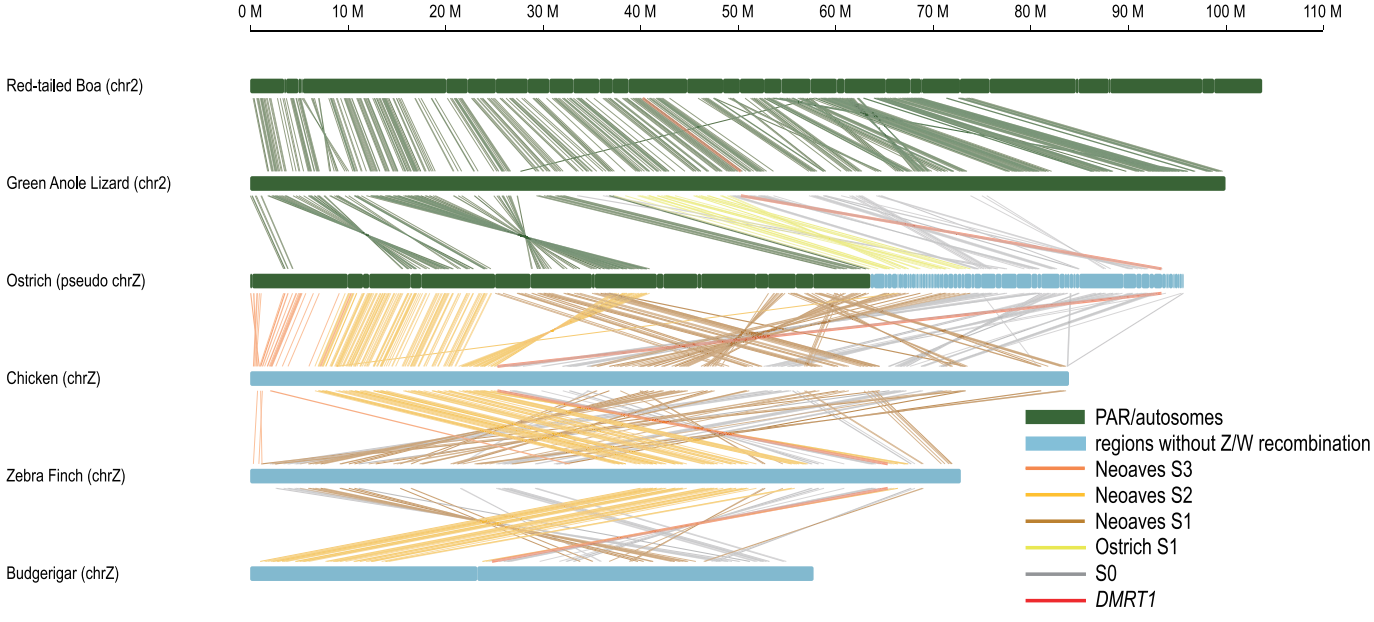


Fig. 3. Gene synteny within avian evolutionary strata. Mapped is the evolutionary strata gene synteny between sequenced reptile (boa and lizard) and avian species. Each line connects a pair of orthologous genes between species, and different colors of lines represent different strata of genes: S0 (gray), ostrich S1 (yellow), Neognathae S1 (brown), Neoaves S2 (orange), *DMRT1* (red). Genomic regions encompassing S0 and S1 have experienced multiple inversions across different species, relocating *DMRT1* to the middle of the Z chromosome after the divergence between Neognathae and Palaeognathae. In contrast, genomic regions within Neoaves S2 and S3 show high level of syntenic conservation across species.

regardless of their different types of sex chromosome inheritance. This matches theoretical predictions that interference effects between linked loci (Hill-Robertson effects), which are ultimately driving Y/W degradation, diminish over evolutionary time owing to the decrease in the number of functional genes that are targets for selection in older strata (*49*).

Discussion

Our results (Fig. 1, summarized in fig. S23) allow us to reconstruct a likely evolutionary formation trajectory of avian sex chromosome strata

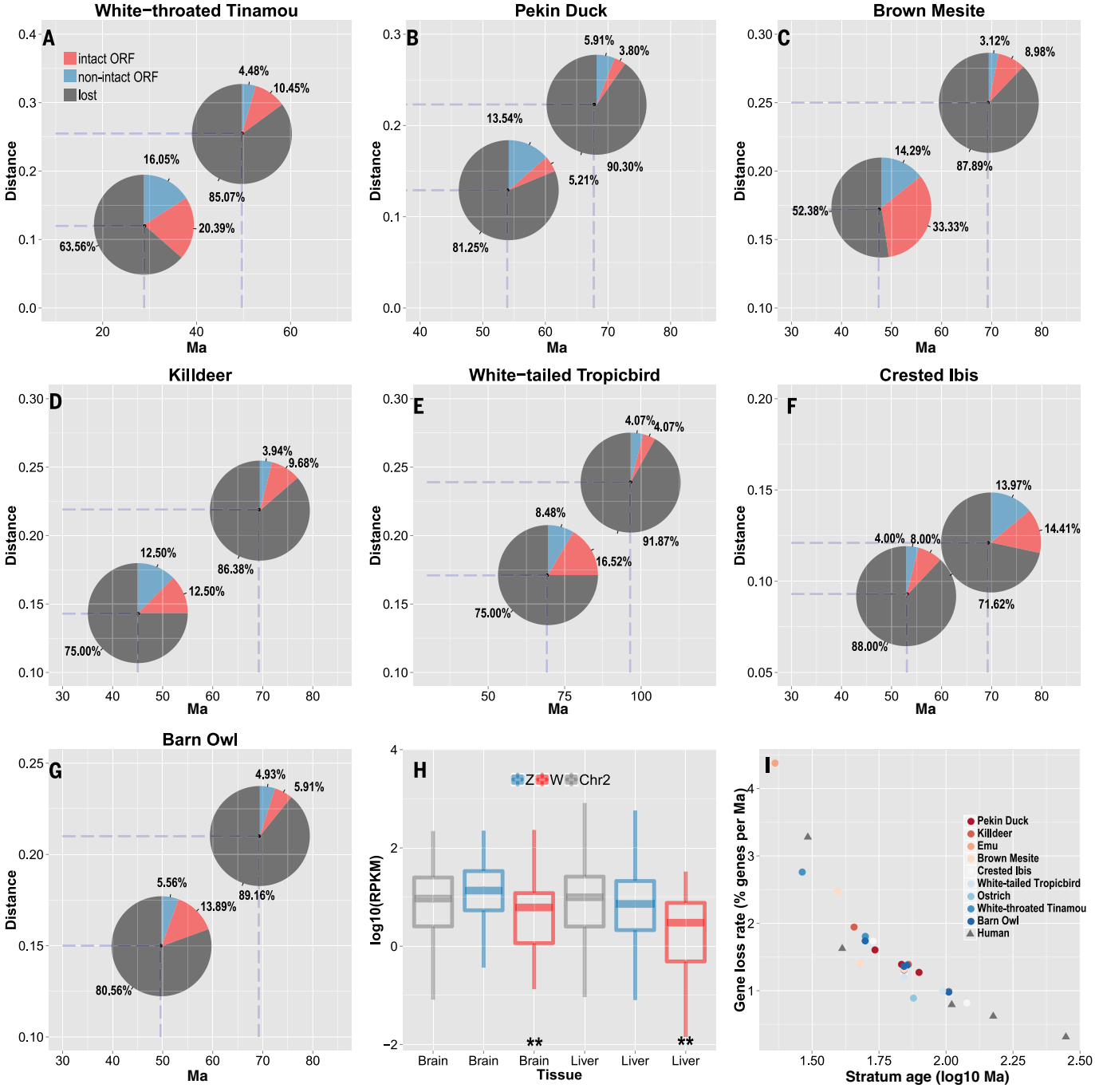


Fig. 4. Evolution of avian W-linked genes. (**A** to **G**) For each stratum of the example species, we show a pie chart of gene composition on the W with regards to whether their open reading frames (ORFs) are predicted to be intact or not. The *x* axis is the estimated age, and *y* axis is the Z/W divergence level of each stratum. Within each pie chart, we show W-linked genes that cannot be found within the assembled sequences (gray), genes with disrupted ORFs that contain premature stop codons or frameshift mutations (blue), and genes with intact ORFs (red). Note that the number of missing genes may be overestimated, owing to the difficulty of assembling W-linked genes surrounded by repetitive sequences. (**H**) Comparison of gene expression between Z and W linked gametologs in ostrich brain and liver samples, relative to their autosomal lizard orthologs. We show significance level of Wilcoxon test comparing the gametolog expression versus that of lizard (**$P < 0.01$). W gametologs show a significant down-regulation of gene expression. (**I**) Correlation between gene loss rate and age of stratum. Each data point represents a stratum of a certain bird species (dot) in Fig. 1 or human (triangle). The data for the human Y chromosome are derived from (*49*, *51*).

(detailed in Fig. 5). We hypothesize that genetic sex determination involving *DMRT1* originated in an ancestor of all birds, and a Z-linked inversion suppressing recombination between chrZ and chrW may have been selected by the accumulation of sexually antagonistic alleles nearby *DMRT1*, forming S0, the only stratum shared among all bird species. This event was followed by numerous independent inversions along the proto–sex chromosome pair accompanying various speciation events. Each event resulted in spreading recombination suppression across genomic regions adjacent to the non-recombining region and decreasing the size of the PAR. Separate evolutionary strata emerged among the different Palaeognathae species. We propose that Neognathae stratum S1 emerged in the ancestor of Neognathae, possibly involving the translocation of the Z-linked *DMRT1* gene toward the center of the Z chromosome. This was followed by the independent formation of stratum S2 in Galliformes, Anseriformes, and the ancestor of Neoaves. Finally some, but not all, Neognathae lineages formed independent S3 strata.

Several general patterns also emerge from comparing the evolutionary history of avian and mammalian sex chromosomes. Unlike in mammals, no sex-autosome fusions contributed to the formation of ancestral avian evolutionary strata, but recombination became abolished in a similar stepwise manner in both birds and mammals. The observation that recombination suppression gradually spreads outwards from the region containing the putative sex-determining gene *DMRT1* (Fig. 1 and fig. S9) in birds mimics patterns of strata formation in mammals, where the oldest stratum contains the sex-determining *SRY* gene. We infer that if inversions caused recombination suppression in birds, they must have occurred on both the Z and W chromosomes (Fig. 5). In contrast, all the inversions suppressing recombination in mammals probably occurred on the Y (*13*).

This different pattern is consistent with a model in which sexual antagonism drives the evolution of suppressed recombination between sex chromosomes (*8*, *50*). Assuming that most sexual antagonism is caused by sexual selection in males, alleles that are beneficial to males but detrimental to females may be more common than female-beneficial male-detrimental mutations (*8*, *50*). In XY systems, Y-linked inversions that trap such male-beneficial female-detrimental alleles into the male-limited region along the Y will be immediately selected for. In ZW systems, however, neither Z- nor W-linked inversions can restrict such alleles to males, but either can be favored by preventing male-beneficial sexual antagonistic alleles from recombining onto the female-specific W chromosome. In the chicken, male function genes are accumulating on the Z chromosome, including a large array of testis-expressed genes termed "Z amplicons" (*3*), providing empirical support for the accumulation of sex-specific genes along the sex chromosomes (table S7).

In conclusion, our results provide a detailed view of the evolutionary history of avian sex chromosomes, involving both shared and lineage-specific recombination suppression events. These events differ by their emergence times, the genomic regions spanned, and the tempo of Z/W differentiation within the affected region, which together underlie the diversity of avian sex chromosome composition. Even within homologous strata of the same age and size, we find marked variation in the extent to which they have lost their W-linked genes in different lineages. It is not entirely clear why rates of avian sex chromosome evolution should differ so dramatically, and whether, or to what extent, this can be attributed to sexual selection, which is widely manifested

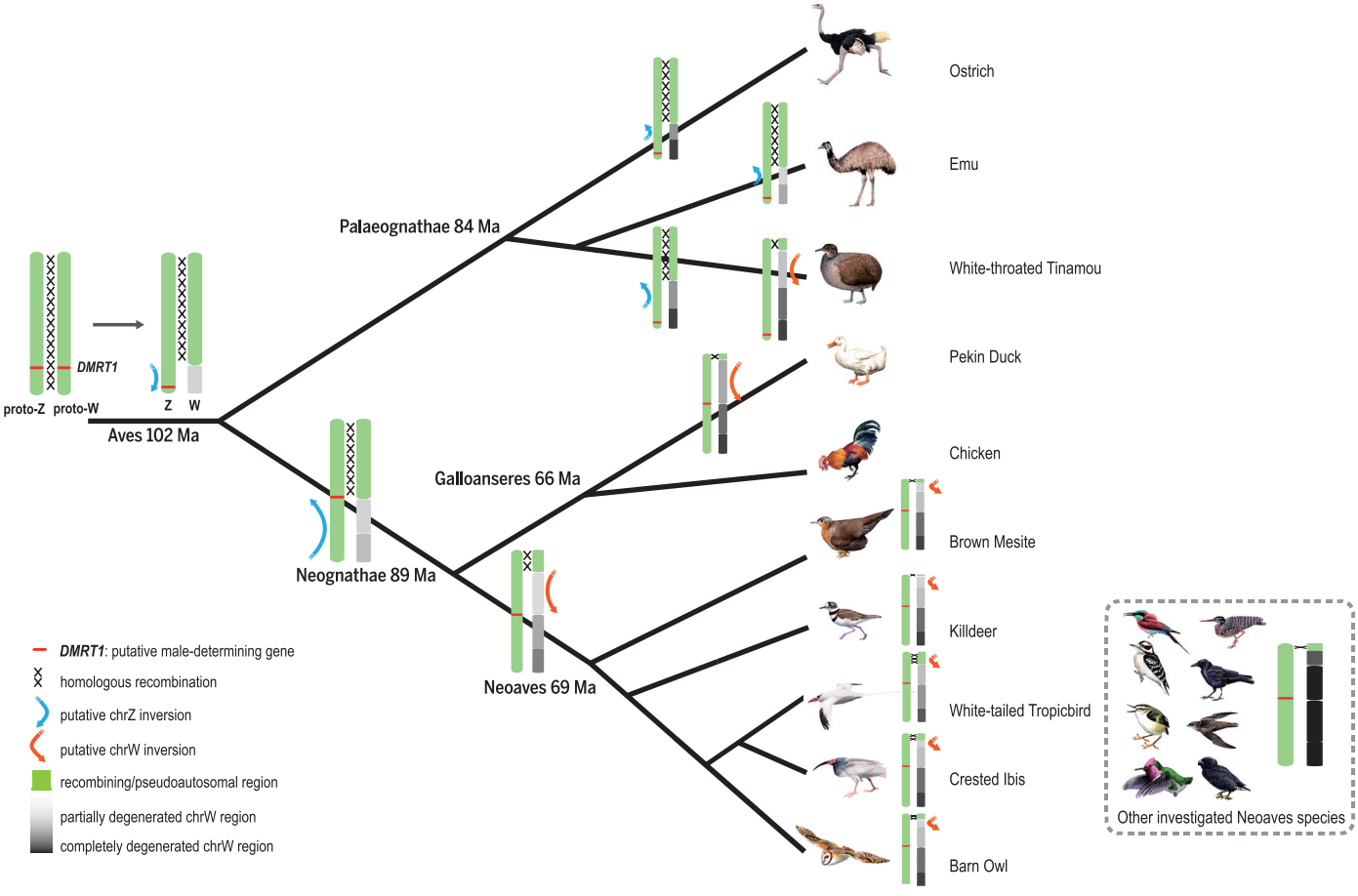


Fig. 5. Hypothesized formation scenarios of avian evolutionary strata. We designate *DMRT1* on the chromosome with the red bar, and homologous recombination between proto–sex chromosomes at PAR (in green) with intersecting lines. Gray or black regions on the W chromosome represent the female-specific region lacking recombination, with darker color representing higher level of differentiation between Z/W. We also show inferred Z- or W-linked inversions that have led to the formation of certain evolutionary strata on the tree.

in birds. Surprisingly, several Neognathae species have retained a large number of genes within the nonrecombining segment of their W chromosome (Table 1), contrary to the general assumption that ancient sex-linked regions invariably degenerate almost completely.

REFERENCE AND NOTES

1. J. J. Bull, *Evolution of Sex Determining Mechanisms* (Benjamin/Cummings, Menlo Park, CA, 1983).
2. A. K. Fridolfsson *et al*., Evolution of the avian sex chromosomes from an ancestral pair of autosomes. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 8147–8152 (1998). doi: 10.1073/pnas.95.14.8147; pmid: 9653155
3. D. W. Bellott *et al*., Convergent evolution of chicken Z and human X chromosomes by expansion and gene acquisition. *Nature* **466**, 612–616 (2010). doi: 10.1038/nature09172; pmid: 20622855
4. D. Cortez *et al*., Origins and functional evolution of Y chromosomes across mammals. *Nature* **508**, 488–493 (2014). doi: 10.1038/nature13151; pmid: 24759410
5. M. T. Ross *et al*., The DNA sequence of the human X chromosome. *Nature* **434**, 325–337 (2005). doi: 10.1038/nature03440; pmid: 15772651
6. B. Charlesworth, D. Charlesworth, The degeneration of Y chromosomes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **355**, 1563–1572 (2000). doi: 10.1098/rstb.2000.0717; pmid: 11127901
7. H. Skaletsky *et al*., The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**, 825–837 (2003). doi: 10.1038/nature01722; pmid: 12815422
8. W. R. Rice, The accumulation of sexually antagonistic genes as a selective agent promoting the evolution of reduced recombination between primitive sex chromosomes. *Evolution* **41**, 911–914 (1987). doi: 10.2307/2408899
9. B. T. Lahn, D. C. Page, Four evolutionary strata on the human X chromosome. *Science* **286**, 964–967 (1999). doi: 10.1126/science.286.5441.964; pmid: 10542153
10. K. Nam, H. Ellegren, The chicken (Gallus gallus) Z chromosome contains at least three nonlinear evolutionary strata. *Genetics* **180**, 1131–1136 (2008). doi: 10.1534/genetics.108.090324; pmid: 18791248
11. G. Li *et al*., Comparative analysis of mammalian Y chromosomes illuminates ancestral structure and lineage-specific evolution. *Genome Res.* **23**, 1486–1495 (2013). doi: 10.1101/gr.154286.112; pmid: 23788650
12. D. W. Bellott *et al*., Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature* **508**, 494–499 (2014). doi: 10.1038/nature13206; pmid: 24759411
13. C. Lemaitre *et al*., Footprints of inversions at present and past pseudoautosomal boundaries in human sex chromosomes. *Genome Biol. Evol.* **1**, 56–66 (2009). doi: 10.1093/gbe/evp006; pmid: 20333177
14. S. A. Sandstedt, P. K. Tucker, Evolutionary strata on the mouse X chromosome correspond to strata on the human X chromosome. *Genome Res.* **14**, 267–272 (2004). doi: 10.1101/gr.1796204; pmid: 14762062
15. K. L. Ayers *et al*., RNA sequencing reveals sexually dimorphic gene expression before gonadal differentiation in chicken and allows comprehensive annotation of the W-chromosome. *Genome Biol.* **14**, R26 (2013). doi: 10.1186/gb-2013-14-3-r26; pmid: 23531366
16. A. E. Wright, H. K. Moghadam, J. E. Mank, Trade-off between selection for dosage compensation and masculinization on the avian Z chromosome. *Genetics* **192**, 1433–1445 (2012). doi: 10.1534/genetics.112.145102; pmid: 22997237
17. L. J. Handley, H. Ceplitis, H. Ellegren, Evolutionary strata on the chicken Z chromosome: Implications for sex chromosome evolution. *Genetics* **167**, 367–376 (2004). doi: 10.1534/genetics.167.1.367; pmid: 15166161
18. H. Ellegren, A. Carmichael, Multiple and independent cessation of recombination between avian sex chromosomes. *Genetics* **158**, 325–331 (2001). pmid: 11333240
19. R. S. de Kloet, S. R. de Kloet, Evolution of the spindlin gene in birds: Independent cessation of the recombination of sex chromosomes at the spindlin locus in neognathous birds and tinamous, a palaeognathous avian family. *Genetica* **119**, 333–342 (2003). doi: 10.1023/B:GENE.0000003842.72339.df; pmid: 14686612
20. J. García-Moreno, D. P. Mindell, Rooting a phylogeny with homologous genes on opposite sex chromosomes (gametologs): A case study using avian CHD. *Mol. Biol. Evol.* **17**, 1826–1832 (2000). doi: 10.1093/oxfordjournals.molbev.a026283; pmid: 11110898
21. J. Rutkowska, M. Lagisz, S. Nakagawa, The long and the short of avian W chromosomes: No evidence for gradual W shortening. *Biol. Lett.* **8**, 636–638 (2012). doi: 10.1098/rsbl.2012.0083; pmid: 22417794
22. S. J. Hackett *et al*., A phylogenomic study of birds reveals their evolutionary history. *Science* **320**, 1763–1768 (2008). doi: 10.1126/science.1157704; pmid: 18583609
23. E. D. Jarvis *et al*., Whole genome analyses resolve the early branches to the tree of life of modern birds. *Science* **346**, 1320–1331 (2014).
24. A. Ogawa, K. Murata, S. Mizuno, The location of Z- and W-linked marker genes and sequence on the homomorphic sex chromosomes of the ostrich and the emu. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 4415–4418 (1998). doi: 10.1073/pnas.95.8.4415; pmid: 9539751
25. C. Nishida-Umehara *et al*., Differentiation of Z and W chromosomes revealed by replication banding and FISH mapping of sex-chromosome-linked DNA markers in the cassowary (Aves, Ratitae). *Chromosome Res.* **7**, 635–640 (1999). doi: 10.1023/A:1009236103013; pmid: 10628664
26. Y. Tsuda, C. Nishida-Umehara, J. Ishijima, K. Yamada, Y. Matsuda, Comparison of the Z and W sex chromosomal architectures in elegant crested tinamou (Eudromia elegans) and ostrich (Struthio camelus) and the process of sex chromosome differentiation in palaeognathous birds. *Chromosoma* **116**, 159–173 (2007). doi: 10.1007/s00412-006-0088-y; pmid: 17219176
27. M. I. Pigozzi, Diverse stages of sex-chromosome differentiation in tinamid birds: Evidence from crossover analysis in *Eudromia elegans* and *Crypturellus tataupa*. *Genetica* **139**, 771–777 (2011). doi: 10.1007/s10709-011-9581-1; pmid: 21567220
28. M. I. Pigozzi, Origin and evolution of the sex chromosomes in birds. *Biocell* **23**, 79–95 (1999). pmid: 10904535
29. G. Zhang *et al*., Comparative genomics across modern bird species reveal insights into avian genome evolution and adaptation. *Science* **346**, 1311–1320 (2014).
30. Y. Huang *et al*., The duck genome and transcriptome provide insight into an avian influenza virus reservoir species. *Nat. Genet.* **45**, 776–783 (2013). doi: 10.1038/ng.2657; pmid: 23749191
31. B. Vicoso, V. B. Kaiser, D. Bachtrog, Sex-biased gene expression at homomorphic sex chromosomes in emus and its implication for sex chromosome evolution. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 6453–6458 (2013). doi: 10.1073/pnas.1217027110; pmid: 23547111
32. L. Ye *et al*., A vertebrate case study of the quality of assemblies derived from next-generation sequences. *Genome Biol.* **12**, R31 (2011). doi: 10.1186/gb-2011-12-3-r31; pmid: 21453517
33. W. L. Fan *et al*., Genome-wide patterns of genetic variation in two domestic chickens. *Genome Biol. Evol.* **5**, 1376–1392 (2013). doi: 10.1093/gbe/evt097; pmid: 23814129
34. Materials and methods are available as supplementary material on *Science* Online.
35. I. Nanda, K. Schlegelmilch, T. Haaf, M. Schartl, M. Schmid, Synteny conservation of the Z chromosome in 14 avian species (11 families) supports a role for Z dosage in avian sex determination. *Cytogenet. Genome Res.* **122**, 150–156 (2008). doi: 10.1159/000163092; pmid: 19096210
36. R. A. Dalloul *et al*., Multi-platform next-generation sequencing of the domestic turkey (Meleagris gallopavo): Genome assembly and analysis. *PLOS Biol.* **8**, e1000475 (2010). doi: 10.1371/journal.pbio.1000475; pmid: 20838655
37. W. C. Warren *et al*., The genome of a songbird. *Nature* **464**, 757–762 (2010). doi: 10.1038/nature08819; pmid: 20360741
38. H. Ellegren *et al*., The genomic landscape of species divergence in Ficedula flycatchers. *Nature* **491**, 756–760 (2012). pmid: 23103876
39. C. A. Smith *et al*., The avian Z-linked gene DMRT1 is required for male sex determination in the chicken. *Nature* **461**, 267–271 (2009). doi: 10.1038/nature08298; pmid: 19710650
40. S. Guioli, R. Lovell-Badge, J. M. Turner, Error-prone ZW pairing and no evidence for meiotic sex chromosome inactivation in the chicken germ line. *PLOS Genet.* **8**, e1002560 (2012). doi: 10.1371/journal.pgen.1002560; pmid: 22412389
41. H. P. Yazdi, H. Ellegren, Old but not (so) degenerated—slow evolution of largely homomorphic sex chromosomes in ratites. *Mol. Biol. Evol.* **31**, 1444–1453 (2014). doi: 10.1093/molbev/msu101; pmid: 24618361
42. J. Alföldi *et al*., The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature* **477**, 587–591 (2011). doi: 10.1038/nature10390; pmid: 21881562
43. K. R. Bradnam *et al*., Assemblathon 2: Evaluating de novo methods of genome assembly in three vertebrate species. *Gigascience* **2**, 10 (2013). doi: 10.1186/2047-217X-2-10; pmid: 23870653
44. A. Kawai *et al*., Different origins of bird and reptile sex chromosomes inferred from comparative mapping of chicken Z-linked genes. *Cytogenet. Genome Res.* **117**, 92–102 (2007). doi: 10.1159/000103169; pmid: 17675849
45. A. Kawai *et al*., The ZW sex chromosomes of Gekko hokouensis (Gekkonidae, Squamata) represent highly conserved homology with those of avian species. *Chromosoma* **118**, 43–51 (2009). doi: 10.1007/s00412-008-0176-2; pmid: 18685858
46. S. Berlin, H. Ellegren, Fast accumulation of nonsynonymous mutations on the female-specific W chromosome in birds. *J. Mol. Evol.* **62**, 66–72 (2006). doi: 10.1007/s00239-005-0067-6; pmid: 16320115
47. Y. Itoh, K. Kampf, A. P. Arnold, Disruption of FEM1C-W gene in zebra finch: Evolutionary insights on avian ZW genes. *Chromosoma* **118**, 323–334 (2009). doi: 10.1007/s00412-008-0199-8; pmid: 19139913
48. J. F. Hughes *et al*., Strict evolutionary conservation followed rapid gene loss on human and rhesus Y chromosomes. *Nature* **483**, 82–86 (2012). doi: 10.1038/nature10843; pmid: 22367542
49. D. Bachtrog, The temporal dynamics of processes underlying Y chromosome degeneration. *Genetics* **179**, 1513–1525 (2008). doi: 10.1534/genetics.107.084012; pmid: 18562655
50. D. Charlesworth, B. Charlesworth, Sex differences in fitness and selection for centric fusions between sex-chromosomes and autosomes. *Genet. Res.* **35**, 205–214 (1980). doi: 10.1017/S0016672300014051; pmid: 6930353
51. M. A. Wilson Sayres, K. D. Makova, Gene survival and death on the human Y chromosome. *Mol. Biol. Evol.* **30**, 781–787 (2013). doi: 10.1093/molbev/mss267; pmid: 23223713

ACKNOWLEDGMENTS

We thank B. Vicoso, R. Bowie, M. Sayres, I. Nanda, and D. Charlesworth for insightful comments and discussions; A. Thais for sharing the green anole lizard photo; O. Ryder and W. Wang for providing the ostrich sample; and the teams of N. Xiao and Z. Bin at OpGen Inc. for performing the optical mapping and technical assistance. We also thank five anonymous reviewers for constructive comments and suggestion. This work was supported by the National Key Basic Research Program of China (973 Program: 2013CB945200) to G.Z. Q.Z. is supported by grants from the National Institutes of Health (GM076007 and GM093182) and a Packard Fellowship to D.B. Raw sequence reads and assembled genomes and gene annotation are deposited at http://phybirds.genomics.org.cn/genome.jsp

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/346/6215/1246338/suppl/DC1
Materials and Methods
Figs. S1 to S23
Tables S1 to S7
References

23 September 2013; accepted 8 August 2014
10.1126/science.1246338

RESEARCH ARTICLE SUMMARY

AVIAN GENOMICS

Convergent transcriptional specializations in the brains of humans and song-learning birds

Andreas R. Pfenning,* Erina Hara, Osceola Whitney, Miriam V. Rivas, Rui Wang, Petra L. Roulhac, Jason T. Howard, Morgan Wirthlin, Peter V. Lovell, Ganeshkumar Ganapathy, Jacquelyn Mouncastle, M. Arthur Moseley, J. Will Thompson, Erik J. Soderblom, Atsushi Iriki, Masaki Kato, M. Thomas P. Gilbert, Guojie Zhang, Trygve Bakken, Angie Bongaarts, Amy Bernard, Ed Lein, Claudio V. Mello, Alexander J. Hartemink,* Erich D. Jarvis*

INTRODUCTION: Vocal learning, the ability to imitate sounds, is a trait that has undergone convergent evolution in several lineages of birds and mammals, including song-learning birds and humans. This behavior requires cortical and striatal vocal brain regions, which form unique connections in vocal-learning species. These regions have been found to have specialized gene expression within some species, but the patterns of specialization across vocal-learning bird and mammal species have not been systematically explored.

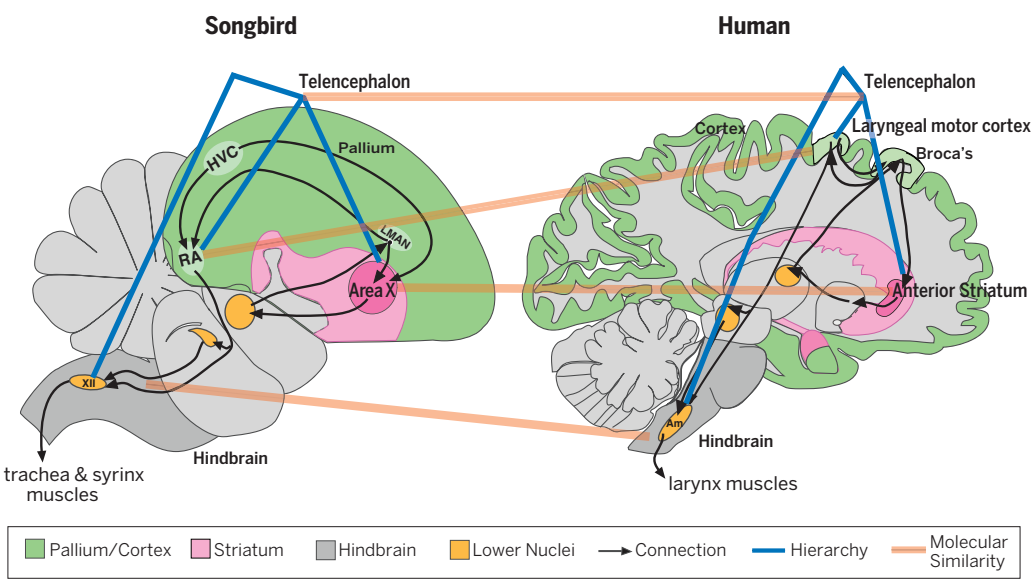
RATIONALE: The sequencing of genomes representing all major vocal-learning and vocal-nonlearning avian lineages has allowed us to develop the genomic tools to measure anatomical gene expression across species. Here, we asked whether behavioral and anatomical convergence is associated with gene expression convergence in the brains of vocal-learning birds and humans.

RESULTS: We developed a computational approach that discovers homologous and convergent specialized anatomical gene expression profiles. This includes generating hierarchically organized gene expression specialization trees for each species and a dynamic programming algorithm that finds the optimal alignment between species brain trees. We applied this approach to brain region gene expression databases of thousands of samples and genes that we and others generated from multiple species, including humans and song-learning birds (songbird, parrot, and hummingbird) as well as vocal-nonlearning nonhuman primates (macaque) and birds (dove and quail). Our results confirmed the recently revised understanding of the relationships between avian and mammalian brains. We further found that songbird Area X, a striatal region necessary for vocal learning, was most similar to a part of the human striatum activated during speech production. The RA (robust nucleus of the arcopallium) analog of song-learning birds, necessary for song production, was most similar to laryngeal motor cortex regions in humans that control speech production. More than 50 genes contributed to their convergent specialization and were enriched in motor control and neural connectivity functions. These patterns were not found in vocal nonlearners, but songbird RA was similar to layer 5 of primate motor cortex for another set of genes, supporting previous hypotheses about the similarity of these cell types between bird and mammal brains.

ON OUR WEB SITE

Read the full article at http://dx.doi.org/10.1126/science.1256846

CONCLUSION: Our approach can accurately and quantitatively identify functionally and molecularly analogous brain regions between species separated by as much as 310 million years from a common ancestor. We were able to identify analogous brain regions for song and speech between birds and humans, and broader homologous brain regions in which these specialized song and speech regions are located, for tens to hundreds of genes. These genes now serve as candidates involved in developing and maintaining the unique connectivity and functional properties of vocal-learning brain circuits shared across species. The finding that convergent neural circuits for vocal learning are accompanied by convergent molecular changes of multiple genes in species separated by millions of years from a common ancestor indicates that brain circuits for complex traits may have limited ways in which they could have evolved from that ancestor. ■



Identifying molecular brain similarities across species. Brain region gene expression specializations were hierarchically organized into specialization trees of each species (blue lines), including for circuits that control learned vocalizations (highlighted green, purple, and orange regions). A set of comparative genomic algorithms found the most similarly specialized regions between songbird and human (orange lines), some of which are convergently evolved.

A complete list of affiliations is available in the full article online.
*Corresponding author. E-mail: apfenning@csail.mit.edu (A.R.P.); amink@cs.duke.edu (A.J.H.); jarvis@neuro.duke.edu (E.D.J.)
Cite this article as A. R. Pfenning *et al.*, *Science* **346**, 1256846 (2014).
DOI: 10.1126/science.1256846

Published by AAAS

Downloaded from www.sciencemag.org on December 14, 2014

RESEARCH ARTICLE

Convergent transcriptional specializations in the brains of humans and song-learning birds

Andreas R. Pfenning,[1]* Erina Hara,[1] Osceola Whitney,[1] Miriam V. Rivas,[1] Rui Wang,[1] Petra L. Roulhac,[1] Jason T. Howard,[1] Morgan Wirthlin,[2] Peter V. Lovell,[2] Ganeshkumar Ganapathy,[1] Jacquelyn Mouncastle,[1] M. Arthur Moseley,[3] J. Will Thompson,[3] Erik J. Soderblom,[3] Atsushi Iriki,[4] Masaki Kato,[4] M. Thomas P. Gilbert,[5,6] Guojie Zhang,[7,8] Trygve Bakken,[9] Angie Bongaarts,[9] Amy Bernard,[9] Ed Lein,[9] Claudio V. Mello,[2] Alexander J. Hartemink,[10]* Erich D. Jarvis[1]*

Song-learning birds and humans share independently evolved similarities in brain pathways for vocal learning that are essential for song and speech and are not found in most other species. Comparisons of brain transcriptomes of song-learning birds and humans relative to vocal nonlearners identified convergent gene expression specializations in specific song and speech brain regions of avian vocal learners and humans. The strongest shared profiles relate bird motor and striatal song-learning nuclei, respectively, with human laryngeal motor cortex and parts of the striatum that control speech production and learning. Most of the associated genes function in motor control and brain connectivity. Thus, convergent behavior and neural connectivity for a complex trait are associated with convergent specialized expression of multiple genes.

Vocal learning is the ability to learn to produce vocalizations by imitating a model. This complex trait convergently evolved in a few lineages of mammals and birds (fig. S1). These include humans, cetaceans (whales and dolphins), pinnepeds (seals and sea lions), bats, and elephants among mammals, and songbirds, parrots, and hummingbirds among birds (*1*). Although nonhuman primates have a limited ability to modify their innate vocalizations, no strong evidence exists that they learn novel vocalizations (*1*–*3*). Vocal-learning species also share the presence of babbling, deafness-induced deterioration of learned vocalizations, dialects, and forebrain circuits that control production and learning of vocalizations (*1*, *4*–*6*). These circuits include a corticostriatal loop and a unique direct connection from motor cortical areas [human laryngeal motor cortex (LMC) and songbird robust nucleus of the arcopallium (RA)] to brainstem vocal motor and respiratory neurons for phonation (Fig. 1).

A potential mechanism for convergent evolution of brain systems is one in which a larger brain structure (superregion) specializes into multiple subregions, each of which performs a distinct function (*7*–*9*). For vocal learning, it has been hypothesized that the superregions of a motor learning or auditory perception pathway have specialized into subregions that control the production of learned vocalizations (*9*–*11*). These functional brain region specializations in song-learning birds (*12*–*17*) and in humans (*18*) are associated with some molecular specializations. One gene, encoding parvalbumin, was found to share specialized expression in brainstem vocal-oral nuclei between song-learning birds and humans (*16*). Some glutamate receptors were found to have specialized expression in forebrain song and speech areas in birds (*19*) and humans (*18*), but a direct comparison between species was not made. The *FOXP2* transcription factor is required for vocal learning in both humans (*20*) and songbirds (*21*–*24*), which suggests that neural circuits for vocal learning across clades may have evolved overlapping molecular mechanisms.

We reasoned that one or more genes underlying a complex trait could show convergent evolution across species, even those separated by tens to hundreds of millions of years from a common ancestor. For example, convergent identical amino acid substitutions have been identified in the opsin gene for light absorption across different vertebrate lineages (*25*); in the melanocorticotropin receptor in mammals and birds for skin and feather color, respectively (*26*); in a gene for yellow wing spots used for courtship in flies (*27*); in Pitx transcription factors for gain and loss of limbs in fish and marine mammals (*28*); in a sodium channel for electric organs in independent lineages of fish (*29*); and in the prestin gene for echolocation in bats and cetaceans (*30*). In addition, for echolocating mammals, genome-scale convergence in amino acid sequence was recently found in many genes involved in hearing (*31*). Likewise, convergent changes in amino acid sequence among vocal-learning birds and mammals have recently been reported (*32*, *33*). However, we are not aware of examples of widespread convergence in gene expression patterns in brain regions involved in convergently evolved behavioral traits.

We examined whether complex behavioral and anatomical convergence is associated with molecular convergence in song-learning birds and humans, species separated from a common ancestor by more than 68 million years (among birds) or 310 million years (between birds and humans) (*34*–*36*). We leveraged the expression profiles of thousands of genes that we (supplementary materials section SM1) and others collected from the brains of vocal-learning and vocal-nonlearning birds (*15*, *16*, *37*), humans (http://human.brain-map.org, June 2013) (*38*), and nonhuman primates (*39*). We developed a hierarchical computational framework to compare brain region specializations across distantly related species.

Framework for finding anatomical molecular similarities across species

To test hypotheses regarding shared molecular specializations between songbird and human brain regions for song and speech control, we collected data from six experiments: (i) 24 total samples from laser-microdissected song control nuclei—Area X, HVC (a letter-based name), LMAN (lateral magnocellular nucleus of the nidopallium), and RA (Fig. 1A)—of three to six adult male zebra finches (*37*), the most commonly studied vocal-learning songbird; (ii) 17 total paired samples from punch biopsy–dissected Area X and adjacent ventral striatum (VS) of male zebra finches hybridized to microarrays of our design (*15*); (iii) 12 total paired samples from laser-microdissected RA analogs and the adjacent motor arcopallium (mAC) in three individuals of each species representing all three vocal-learning avian lineages (songbird, parrot, and hummingbird) and two nonvocal mAC regions (*9*) of two vocal-nonlearning lineages (dove and quail) (*16*), generated for this project (SM1); (iv) from these same animals, laser-microdissected brainstem vocal (nXIIts; 12th tracheosynringeal nucleus) and neck (SSp; supraspinal nucleus) motor neurons, which are found in all vertebrate species; (v) 3702 samples dissected from 231 subregions (table S1A) of the much larger postmortem human brains of six subjects from the Allen Human Brain Atlas (*38*); and (vi) 258 laser-captured samples from different

[1]Department of Neurobiology, Howard Hughes Medical Institute, and Duke University Medical Center, Durham, NC 27710, USA. [2]Department of Behavioral Neuroscience, Oregon Health & Science University, Portland, OR 97239, USA. [3]Duke Proteomics and Metabolomics Core Facility, Center for Genomic and Computational Biology, Duke University Medical Center, Durham, NC 27710, USA. [4]Laboratory for Symbolic Cognitive Development, Brain Science Institute, RIKEN, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan. [5]Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, 1350 Copenhagen, Denmark. [6]Trace and Environmental DNA Laboratory, Department of Environment and Agriculture, Curtin University, Perth, Western Australia 6102, Australia. [7]China National GeneBank, BGI-Shenzhen, Shenzhen 518083, China. [8]Centre for Social Evolution, Department of Biology, University of Copenhagen, DK-2100 Copenhagen, Denmark. [9]Allen Institute for Brain Science, Seattle, WA 98103, USA. [10]Department of Computer Science, Duke University, Durham, NC 27708, USA.
*Corresponding author. E-mail: apfenning@csail.mit.edu (A.R.P.); amink@cs.duke.edu (A.J.H.); jarvis@neuro.duke.edu (E.D.J.)

cortical layers of 12 regions of four rhesus macaques, including major motor and sensory cortices (*39*).

The songbird RA and HVC are part of a vocal motor pathway that controls the production of learned vocalizations, whereas Area X and LMAN are part of a pathway that controls imitation and exploration of vocalizations (Fig. 1A) (*40*, *41*). We used only animals that did not sing after an overnight period of silence alone, because neural activity associated with singing causes changes in expression of thousands of genes in the song nuclei (*37*, *42*). We do not know the premortem vocalizing state of the humans, but did not expect this to prevent us from identifying convergent gene expression specializations as long as one species was silent (the birds), the genes are not sensitive to immediate vocalizing behavior, and the human samples were collected 10 to 30 hours postmortem, when many of the vocalizing-regulated gene expression changes have subsided. The avian samples were profiled on our custom songbird oligonucleotide microarray [zebra finch Agilent oligoarray v2.2 (*37*)]; human samples were profiled on a human microarray (Agilent custom 8×60K array) containing 7473 orthologous genes we identified between human and finch (table S2 and SM2 to SM4).

Because brain regions of different species may show convergent specializations for different sets of genes relative to homologous surrounding brain regions (or for the same genes within more divergent homologous surrounding brain regions), we developed a computational framework to accommodate these and other hypotheses (SM5). First, an anatomical gene expression specialization tree was created for each species to hierarchically organize the expression profiles of all genes from more global to specific regions (e.g., forebrain > cortex > motor cortex > subregion of motor cortex). Although the hierarchy is based on anatomy and not gene expression, the two

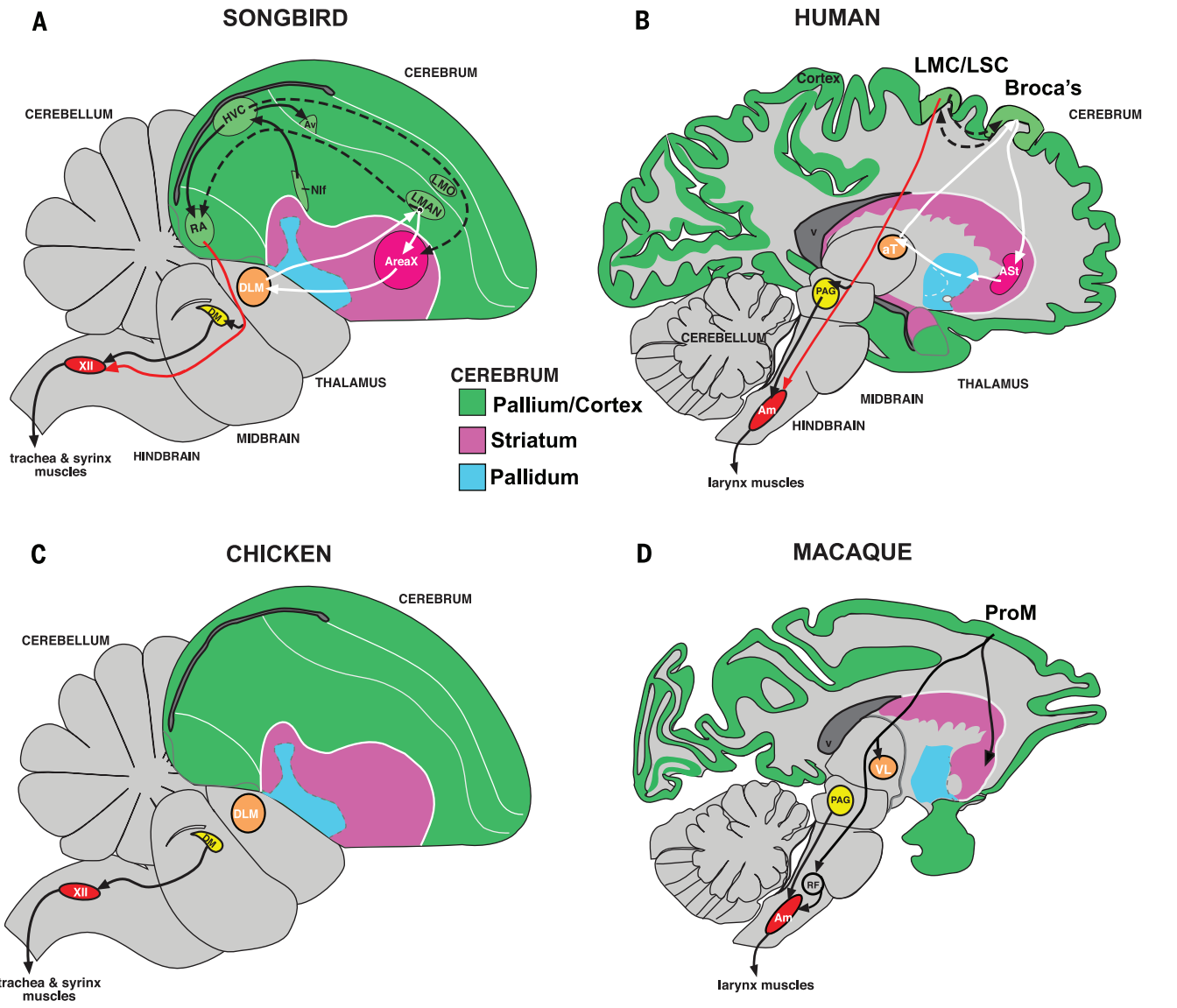


Fig. 1. Comparative brain relationships, connectivity, and cell types among vocal learners and nonlearners. (**A**) Drawing of a zebra finch male brain section showing profiled song nuclei: Area X, HVC, LMAN, RA, and the 12th motor nucleus (XII). (**B**) Drawing of a human brain section showing proposed vocal pathway connectivity including LMC/LSC in the precentral gyrus. Black arrows, connections and regions of the posterior vocal motor pathway; white arrows, connections and regions of the anterior vocal pathway; dashed arrows, connections between the two pathways. The thick blue arrows show the analogous brain regions predicted by this study across finch and human. Red arrows show the direct projections found only in vocal learners, from vocal motor cortex regions to brainstem vocal motor neurons. (**C**) Known connectivity of a vocal-nonlearning bird showing absence of forebrain song nuclei. (**D**) Known connectivity of vocal-nonlearning primates (i.e., macaque) showing presence of forebrain regions that have an indirect projection to nucleus ambiguus (Amb) but have no known role in production of vocalizations.

correspond relatively well, especially at higher levels (Fig. 2A and fig. S2). A node in the tree is a vector of expression levels of all genes for each brain sample or higher grouping of samples (regions); a branch is the difference (specialization) in expression of all genes between two nodes. Second, we calculated the pairwise distances and associated Pearson correlation of all nodes and branches between two species.

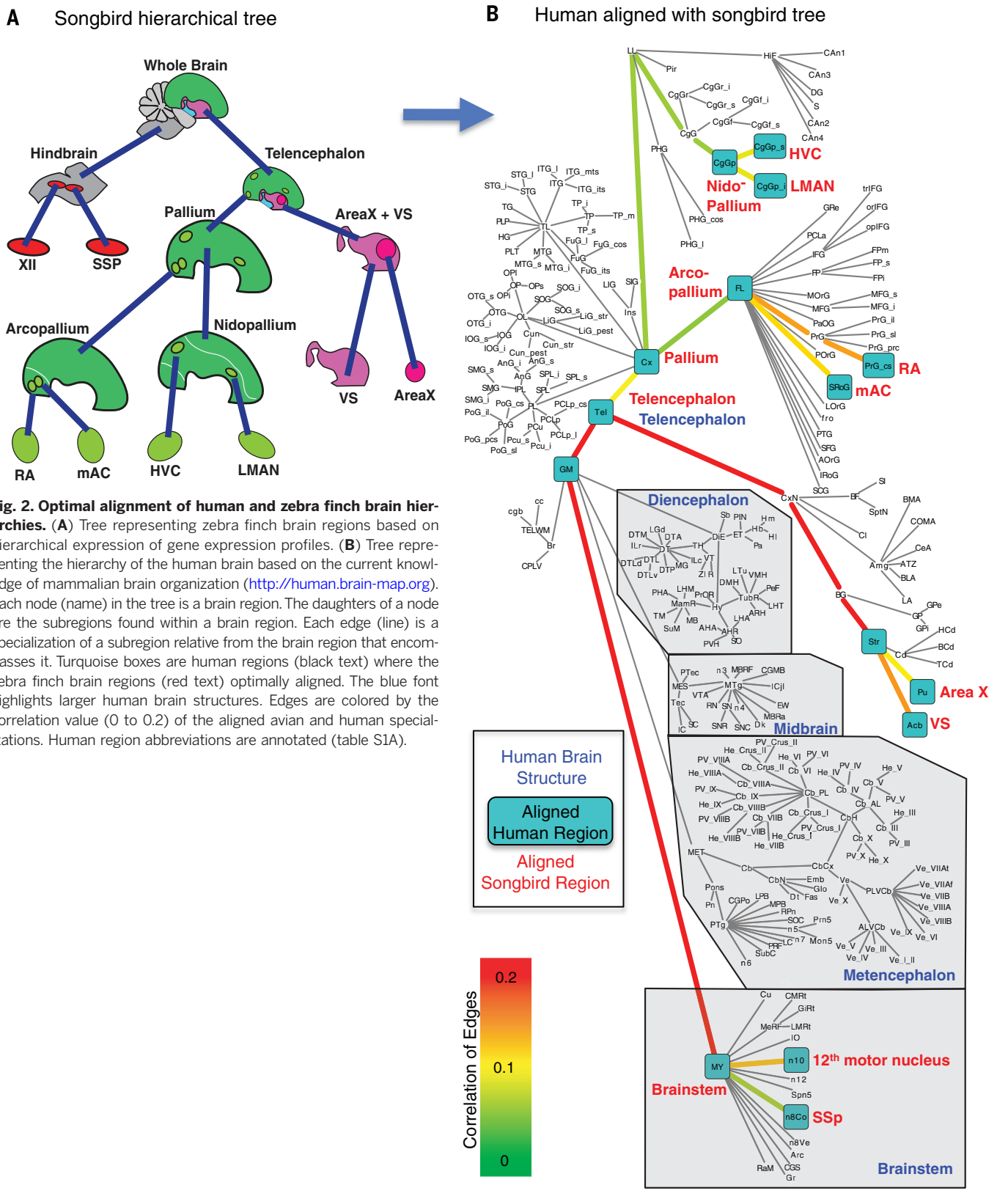


Fig. 2. Optimal alignment of human and zebra finch brain hierarchies. (**A**) Tree representing zebra finch brain regions based on hierarchical expression of gene expression profiles. (**B**) Tree representing the hierarchy of the human brain based on the current knowledge of mammalian brain organization (http://human.brain-map.org). Each node (name) in the tree is a brain region. The daughters of a node are the subregions found within a brain region. Each edge (line) is a specialization of a subregion relative from the brain region that encompasses it. Turquoise boxes are human regions (black text) where the zebra finch brain regions (red text) optimally aligned. The blue font highlights larger human brain structures. Edges are colored by the correlation value (0 to 0.2) of the aligned avian and human specializations. Human region abbreviations are annotated (table S1A).

Third, these distance values were given as input to a dynamic programming algorithm we developed that finds the optimal gene expression specialization alignment of one species' anatomical expression tree (e.g., avian) with another's (e.g., human) while preserving the hierarchical structure of each tree (SM6).

Similarities in gene expression specializations may be due to homology (common origin) or convergence, or may result from shared expression profiles of one or a few cell types within brain regions between species. We found that the Pearson correlations between specializations were low, even in biological replicates, because of noise and large numbers of genes that are not specialized in expression. For example, the mean correlation between biological replicates of zebra finch RA relative to the adjacent mAC specialization is 0.22. Values above 0.09 were typically statistically significant (see below).

Molecular specializations shared between songbird and human brains

We first compared the gene expression profiles between songbird (zebra finch) and human, for which we had the most samples. The dynamic programming algorithm found an optimal alignment between the brain gene expression specialization trees of songbird (Fig. 2A) and human (Fig. 2B; see also Table 1). At high levels of the alignment, the songbird telencephalic regions corresponded most closely to the human telencephalon (Tel) and the songbird brainstem nuclei corresponded to human brainstem nuclei. Within the telencephalon, the songbird pallial regions corresponded most closely to the human cortex (Cx), and the songbird striatal regions corresponded to the human striatum. The songbird pallial regions sampled did not have a strong similarity to the human claustrum (Cl) and had an inverse gene expression relationship with the amygdala (Amg), meaning that the specialization was in the opposite direction (Table 1).

These results support the cortex hypothesis (*33*, *43*, *44*) and contradict the claustrum-amygdala hypothesis of homologies between avian and mammalian pallial brain regions (*33*, *45*). Although some avian striatal regions (Area X and VS) are known to have a sparse population of pallidum-like neurons (*46*–*48*), we did not find a strong correlation with the human pallidum (Table 1), supporting the conclusion that the avian regions are predominantly striatal (*33*, *44*). These findings show that our approach works, in that it recapitulates the consensus comparative anatomy view of avian and mammalian brains; in addition, it provides evidence of competing hypotheses concerning more specific homologies.

At lower levels of the tree alignment (individual brain regions), songbird Area X best aligned to the human putamen (Pu) (Fig. 2B and Table 1), supporting a prediction made on the basis of other evidence (*41*, *44*, *48*). The avian ventral striatum (VS) adjacent to Area X that does not participate in vocal learning best aligned to the human ventral striatum [i.e., nucleus accumbens (Acb) ventral to the putamen] (Fig. 2B). Songbird RA best aligned to the human central sulcus (PrG_cs) within the primary motor cortex [i.e., within the precentral gyrus (PrG)] (Fig. 2B), supporting a prediction of the cortex hypothesis (*4*, *33*, *41*, *49*); the human central sulcus and songbird RA have some of the most robust direct

Table 1. Support and rejection for hypothesized relationships between human and zebra finch brain regions. We allowed Wernicke's area to be any subregion in the superior temporal gyrus; Broca's area to be either the opercular part or the triangular part of the inferior frontal gyrus; the supplementary motor area to be any part of the encompassing superior frontal gyrus. The specialization of the finch forebrain to Area X + VS and finch forebrain to pallium was compared to the human specializations of the amygdala, claustrum, cortex, globus pallidus, and striatum. The finch striatum (Area X + VS) to Area X was compared to caudate, putamen, and nucleus accumbens. The finch specialization of the pallium to RA, HVC, and LMAN was compared to all subregions of the human cortex relative to the entire cortex, as well as each subregion/cortical layer combination in macaque. Significantly supported hypotheses are in bold.

Finch specialization	Human region/macaque cortex layer	Reference	Correlation	Permutation		Abbreviation
				P	FWER	
Forebrain to pallium	**Telencephalon to cortex**	(*4*, *41*)	0.105	0.068	0.218	Cx
	Telencephalon to claustrum	(*45*)	0.007	0.537	0.595	Cl
	Telencephalon to amygdala	(*45*)	−0.083	0.840	1.000	Amg
Forebrain to Area X + VS	**Telencephalon to striatum**	(*41*, *44*)	0.207	0.001	0.001	Str
	Telencephalon to globus pallidus	(*44*, *46*, *48*)	−0.068	0.800	1.000	GP
Striatum to Area X	**Striatum to putamen**	(*41*, *44*)	0.108	0.031	0.082	Pu
	Striatum to caudate	(*41*, *44*)	0.037	0.001	0.393	Cd (BCd)
	Striatum to nucleus accumbens	(*48*)	−0.141	0.995	1.000	Acb
Pallium to RA	**Precentral gyrus, central sulcus**	(*4*, *41*, *49*, *84*)	0.139	0.003	0.017	PrG_cs
	Layer 5, motor cortex	(*44*)	0.132	0.000	0.011	PrGxL5
	Cingulate cortex	(*4*)	0.089	0.000	0.051	CgGf (CgGf_i)
Pallium to HVC	Wernicke's area	(*4*)	0.045	0.302	0.999	STG
	Broca's area	(*4*, *84*)	0.006	0.505	1.000	opIFG or trIFG
	Supplementary motor area	(*4*)	−0.004	0.491	1.000	SFG
	Layers 2/3	(*85*)	0.046	0.208	0.908	PoGxL2
	Layer 4	(*68*)	0.026	0.469	0.998	TAxL4
Pallium to LMAN	Broca's area	(*41*)	0.036	0.148	0.997	opIFG or trIFG
	Pre-SMA	(*41*)	−0.010	1.000	0.556	SFG
	Anterior cingulate	(*84*)	−0.085	1.000	0.849	CgGf
	Layer 4	(*68*)	0.089	0.070	0.249	PoGxL4
Brainstem to XII	Myencephalon to hypoglossal nucleus	(*44*)	0.009	0.440	1.000	n12
	Myencephalon to vagus motor nucleus	No proposal	0.125	0.040	0.040	n10
	Myencephalon to nucleus ambiguus	(*54*)	NA	NA	NA	NA

projections to brainstem motor neurons known (Fig. 1) (*41*, *50*). The mAC lateral to RA, which does not participate in song production (*9*, *51*) and does not make direct projections (*52*, *53*), aligned weakly to the human superior rostral gyrus (SRoG) of the frontal lobe (Fig. 2B). Both HVC and LMAN aligned best to the parietal part of the cingulate cortex, but the correlations were low (Fig. 2B). None of the regions hypothesized to be similar to HVC or LMAN had a strong secondary match (correlations > 0.09), but of the hypothesized regions, the highest correlation values were HVC with Wernicke's auditory speech processing region and LMAN with Broca's speech production and processing region (Table 1). Within the brainstem, the songbird vocal (nXIIts) and neck (SSp) motor neurons could not be aligned to their proposed human homologs—the nucleus ambiguus and supraspinal nucleus, respectively (*16*, *44*, *49*, *54*)—because they were not isolated as part of the Allen Human Brain Atlas data set. However, the closest match for nXIIts in the available human data was to the vagus motor neurons (n10) (Fig. 2B and Table 1). These findings support a number of predictions and enable new predictions about molecular similarities between avian and human brain regions.

Confirming brain region specializations between human and songbird

The dynamic programming algorithm provided an unbiased global alignment between human and songbird brain regions, and the input tree distance measures provided correlation values. However, the statistical significance obtained from the Pearson correlation calculations might be artificially inflated if large numbers of genes have correlated expression. Thus, to estimate the significance of the similarities and perform an independent statistical test, we (i) computed a more stringent permutation *P* value (0.05 threshold) and associated familywise error rate (FWER, 0.2 threshold) for each proposed relationship on the basis of the Pearson correlations between species (SM7), and (ii) performed a linear regression on the nodes of the tree to determine the number of significantly specialized genes in common (SM8), relative to what would be expected by chance (SM9).

Both the permutation and gene number methods gave moderate ($P \approx 0.05$) to strong ($P < 0.001$) statistical support for most zebra finch and human regions aligned by the dynamic programming algorithm at Pearson correlations of ~0.09 and above (Fig. 2B, Table 1, and fig. S3). In particular, songbird striatal regions (Area X and VS) were confirmed by both methods to have the most similar specialized gene expression to the human striatum (Fig. 3A and fig. S3A; Table 1 and table S3A). The avian pallial regions were confirmed to be most similar to the human cortex, with a lower significance by the permutation analyses ($P = 0.068$) and corresponding FWER analyses (0.218) but with the highest number of specialized genes in common of all regions compared (Fig. 3B and fig. S3B; Table 1 and table S3B). The lower significance could possibly reflect greater differences between the avian pallium and human cortex than between the avian and human striatum (*33*, *47*).

Among the striatal regions, both tests confirmed Area X to have the strongest similarity with the putamen (Fig. 3C and fig. S3C; Table 1 and table S3C) but a significant secondary relationship with the adjacent body of the caudate, consistent with the hypothesis that Area X is similar to adjacent parts of the caudate and putamen (*41*). Among the pallial regions, both tests confirmed RA to have the strongest similarity to the central sulcus part of the human primary motor cortex (PrG_cs) but also to the adjacent primary somatosensory part of the central sulcus (PoG_cs) and to a region where the ventral LMC is located (PrG_sl) (Fig. 3D and fig. S3D; Table 1 and table S3D). The similarity to the somatosensory cortex is not too surprising, considering the underappreciated overlap of cells with motor and somatosensory functions in both walls (PrG_cs and PoG_cs) of the central sulcus (*55*). This is supported by a strong similarity in gene expression between PrG and PoG samples (fig. S2C). Another region proposed to be functionally analogous to RA, the cingulate gyrus (*4*), also had strong similarity from both tests (Fig. 3D and fig. S3D; Table 1 and table S3D). For HVC and LMAN, of the hypothesized human brain regions, again the highest similarities were to Wernicke's area (plus Heschl's and transverse auditory gyri; HG, TG) and Broca's area, respectively, but the relationships were not significant according to the permutation and FWER tests (Table 1, fig. S3F, and table S3F). These results reveal that similarities identified by the dynamic programming algorithm are significant, and further reveal more refined similarities with other brain regions according to the anatomical annotations of the Allen Human Brain Atlas.

Zebra finch Area X and RA are most similar to human speech brain regions

The above analyses gave equal weight to samples with the same anatomical annotation, ignoring possible variation that might exist across samples or samples annotated differently according to other hypotheses of mammalian or avian brain organization. To address these concerns, we examined the correlation between the specialization of a finch brain region and each of the individual human brain samples within a matching region of interest, overlaid with the coordinates of brain regions activated during speech production (SM10) (*56*–*62*).

We found that the zebra finch Area X specialization best matched specific human putamen and caudate samples that overlapped and extended with the coordinates of brain regions activated during speech production (Fig. 4A) (*56*–*58*). Similarly, the zebra finch RA specialization best matched specific samples at or near coordinates of a recently hypothesized dorsal LMC (dLMC) and adjacent somatosensory laryngeal cortex (dLSC) (*62*, *63*) within the PrG and PoG, respectively, as well as to vLMC samples within PrG, all regions that have the highest functional activation during speech production (Fig. 4B) (*59*). Moreover, several of the human samples with the highest molecular scores shared with songbird RA, which controls the syrinx and not the tongue, beak, or jaw (*41*), are in proximity to the center of a region activated during the production of vowels (generated by the LMC connections to the larynx), as opposed to the production of tongue, lip, or jaw movements (Fig. 4B) (*59*). The results suggest that the match of RA to LMC/dLSC contributed to the relationship of RA with the broader human primary motor cortex.

All vocal-learning birds and humans show convergent gene expression specializations

A case for convergent specialized gene regulation in vocal-learning brain regions as a general principle would be supported if other avian vocal-learning lineages showed similar convergent specializations with humans, and negative control vocal-nonlearning birds and vocal-nonlearning primates did not. We first looked for possible specializations in the gene expression profiles of the RA analog from species representing the two other vocal-learning lineages (parrots and hummingbirds) compared to the neighboring arcopallium and comparably located regions of the arcopallium of species representing two experimentally determined vocal-nonlearning lineages (dove and quail) (*41*, *64*, *65*). Because the avian samples were run on the zebra finch Agilent microarray, we had to develop a strict filtering pipeline to include only oligonucleotides that mapped to and hybridized to avian genomic DNA across species, yielding a total of 3044 genes each represented by at least one oligonucleotide also present on the human Agilent microarray (SM3).

This reduced set of genes applied to our dynamic programming algorithm recapitulated the strong relationship between gene expression specializations in zebra finch RA and human PrG_cs (fig. S4A). This relationship was stronger for the RA analog of all three vocal learners (fig. S4, A to C) relative to the similarly located central arcopallium of both vocal nonlearners (fig. S4, D and E). The correlations increased for the vocal-learning avian species, but not for the vocal-nonlearning avian species, when comparisons were made to the specific human putative LMC (dLMC and vLMC) and dLSC samples within the PrG_cs and PoG_cs relative to the cortex (fig. S4C, triangles and squares); the correlations were still present when the comparison was relative to only the remaining PrG and PoG (table S4, A to C). Likewise, the RA analogs of all three vocal learners had many more genes specialized in common with the human LMC/dLSC than did the arcopallium of vocal nonlearners, and the number of specialized genes in common was significant only for the vocal learners (Fig. 3, E and F, and table S4, B and C). For a control analysis, we identified the common specialized genes between all the arcopallium samples of each avian species and the entire human frontal lobe, and found them to be similar across all avian species (Fig. 3G).

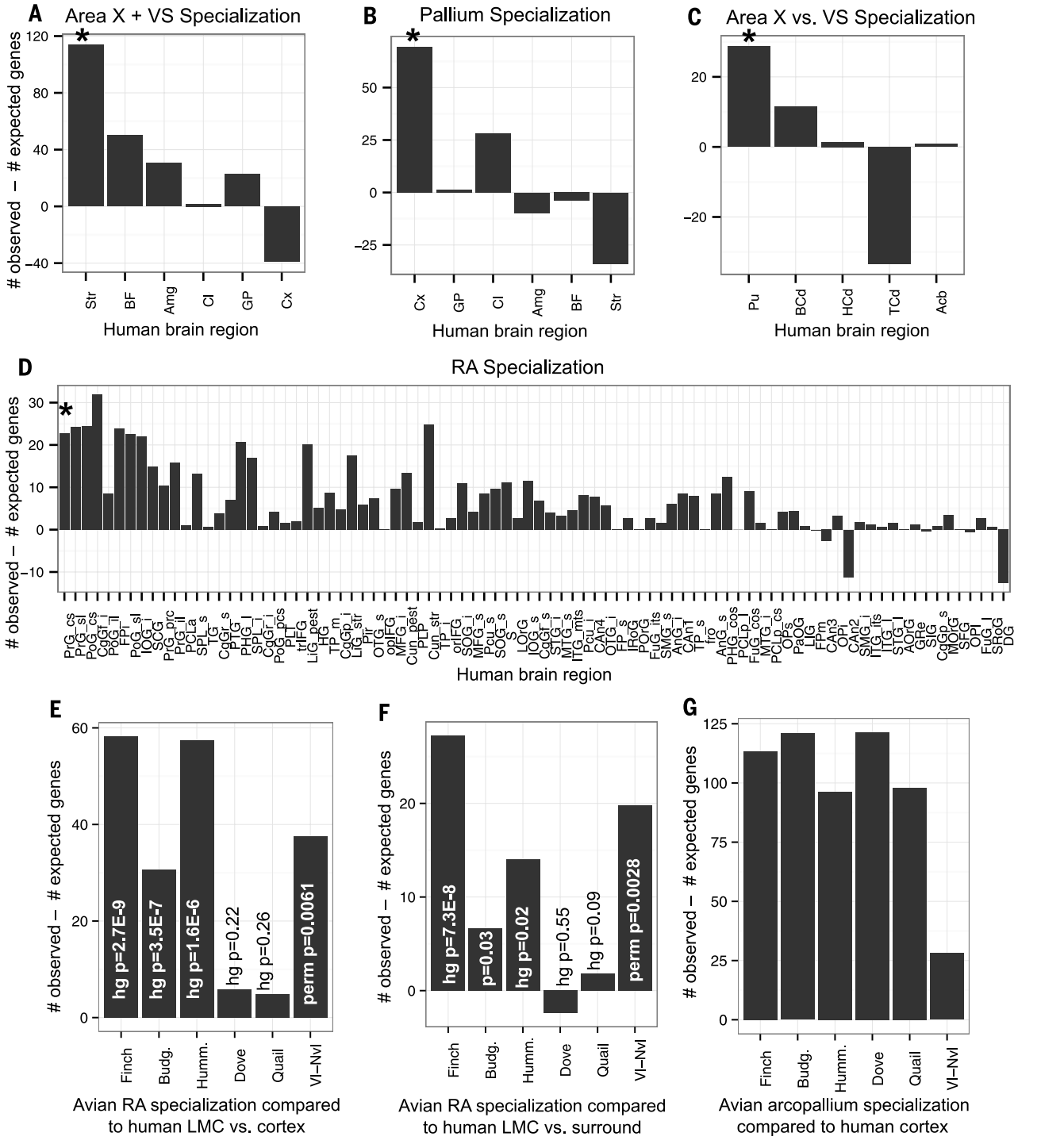


Fig. 3. Relative number of genes with significantly shared specialized expression between avian and human brain regions. Each panel shows a plot of the number of genes significantly specialized ($P < 0.05$; hypergeometric test) in common between the avian and human samples relative to the number of genes expected to be specialized by chance. (**A**) Finch Area X + VS specialization compared to all subregions of the human telencephalon. (**B**) Finch pallial region (RA, neighboring arcopallium, HVC, and LMAN combined) specialization compared to all subregions of the human telencephalon. (**C**) Finch Area X specialization compared to all subregions of the human striatum. (**D**) Finch RA specialization compared to the specialization of every subregion from the human cortex, which optimally aligned to the zebra finch pallium. (**E**) Avian RA analogs (vocal learners) and mAC (nonlearners) relative to the adjacent arcopallium compared to human LMC/dLSC relative to cortex. (**F**) Avian RA analogs (vocal learners) and mAC (nonlearners) relative to the arcopallium compared to human LMC/dLSC relative to PrG/PoG. (**G**) Avian arcopallium versus whole brain specialized genes compared to human cortex versus whole brain specialized genes. In (A) to (D), asterisks denote the human specializations determined to be similar to the avian specialization on the basis of the optimal alignment and correlation. In (E) to (G), P values less than 0.05 indicate that the number of specialized genes is greater than chance according to a hypergeometric test.

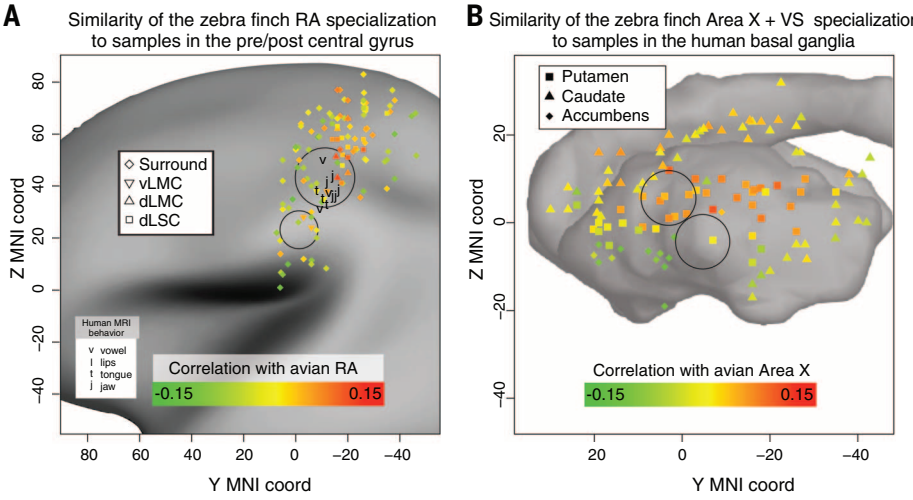


Fig. 4. Overlap of regions with convergent specialized expression and speech activation. Shown are the coordinates in MNI (Montreal Neurological Institute) space of each human brain microarray sample from the striatum (**A**) and the precentral/postcentral gyrus (**B**). The different subregions within the striatum and precentral/postcentral gyrus are labeled by differently shaped symbols. These points are placed on top of a representative image of the cortical surface map from the Allen Human Brain Atlas (*38*). Only the left hemisphere, which has higher sample density, is shown. In (A), each point is colored on the basis of the correlation between the specialization of that sample relative to the entire human striatum, and of Area X relative to Area X and VS together. Circles represent high-confidence regions of speech activation from multiple coordinates in multiple studies (*56*–*62*). In (B), each point is colored on the basis of the correlation between the specialization of that sample relative to the entire human PrG/PoG, and of zebra finch's RA specialization. A similar plot with the average of all vocal learners is shown in fig. S4C.

To test whether the same set of genes led to these independent patterns of convergent evolution, we extended our linear regression model to identify genes that were differentially specialized with humans in all three vocal-learning avian species relative to two vocal-nonlearning avian species (SM9). This vector of genes and expression levels in all vocal-learning birds had a positive correlation with genes specialized in LMC/dLSC relative to the surrounding PrG and PoG (0.14, permutation $P = 0.016$). A significant number of genes overlapped between the human LMC/dLSC and the vocal-learning minus vocal-nonlearning birds (Fig. 3, E and F, permutation test; Table 1). Overall, these findings indicate that the avian RA analogs of each vocal-learning lineage have mostly the same convergent molecular specializations with human LMC/dLSC.

Songbird RA shows greater similarity to human LMC than to nonhuman primate motor cortex

We conducted the macaque comparisons separately because the macaque microarray data were from more limited regions (mainly in the cortex), contained different cell layers of each cortical region (as opposed to all cortical layers combined in each human sample), and had fewer orthologous genes (4702) with zebra finch than with human (7473). All of our approaches (correlation, gene number, and permutation) showed that songbird RA had the highest match to layer 5 cells of macaque primary motor cortex (with a secondary match to layer 5 of premotor cortex) relative to all other layers within the primary motor cortex and to all layers of all other cortical regions sampled (cingulate, auditory, visual, and other cortices; Table 1, fig. S5, A and B, and tables S3G and S4D). HVC had the best match, although weaker, to macaque layer 2 of the PoG and other cortical regions (Table 1, fig. S5, C and D, and table S3H). Both findings are consistent with the cortex hypothesis of homologies (*41*, *66*, *67*). However, LMAN, also part of the nidopallium, had the best match to layer 4 of the PoG and other cortical regions, contradicting previous hypotheses (*41*, *67*) (Table 1, fig. S5, E and F, and table S3I) but consistent with a recent study comparing chicken nidopallium to mouse cortical layers (*68*).

Comparing the arcopallium samples from each of the avian species with the specializations of human LMC/dLSC versus macaque layer 5, we found a weak, nonsignificant difference in the correlation between the arcopallium regions of avian vocal learners versus nonlearners and macaque layer 5 cells (table S4D), but a stronger significant correlation between the RA analogs of avian vocal learners versus nonlearners and human LMC/dLSC versus macaque layer 5 cells (table S4E). These findings demonstrate the similarity of avian RA and arcopallium to mammalian layer 5 cells, but a greater similarity of each vocal-learning species' RA analog to human laryngeal cortex regions (which also contain layer 5 cells) relative to nonhuman primate motor cortex.

Function of genes specialized between avian and human brains

We conducted a linear regression analysis to identify the specific genes with statistically significant differences for each species (SM8) and performed functional enrichment analyses relative to annotated pathways (SM11). We found that 288 genes contributed significantly to the shared specialization between the avian and human striatum (table S3A). These included those with positive control in situ hybridization data showing higher expression that distinguishes the striatum from other regions in both birds and mammals (fig. S6, A and B, and table S3A) (*43*). Functional enrichment analysis of these genes found movement disorders (68 genes) and disorders of the basal ganglia (55 genes) as two of the top categories (table S5A), consistent with known functions of the striatum. Similarly, 342 genes contributed to the shared specialization between avian pallium and human cortex, and also included those validated by in situ hybridization (fig. S6, C and D, and table S3B) (*43*). The top enrichment categories were behavior (64 genes) and neurotransmission (38 genes; table S5B). These findings show that our linear regression approach correctly identifies genes contributing to shared specializations between avian and mammalian brains.

We identified 78 genes that contributed significantly to the shared specialized expression between songbird Area X and human putamen (fig. S7 and table S3C). In situ hybridization analyses of 13 randomly selected genes revealed that those with a permutation $P < 0.015$ had a true positive rate of ~90% for differential expression in Area X (fig. S7, highlighted genes, fig. S8, and table S3C). The top enrichment categories were movement disorders (21 genes) and ataxia (8 genes; table S5C), consistent with motor-driven neural activity and gene regulation in Area X (*4*, *42*), and with functional magnetic resonance activation during speech production in the matching human striatal regions (*57*).

We identified 40 genes that contributed to the shared specialized expression between songbird RA and human PrG_cs (table S3D); a subset of these (10 genes) contributed ~50% to the shared specialization between songbird RA and macaque layer 5 of motor cortex (table S3D). In contrast, we found a mostly nonoverlapping set of 55 genes (only 3 overlapping) that contributed to the convergent shared specialized expression between the RA analogs of vocal learners and human LMC/dLSC (Fig. 5 and table S6). These findings indicate that the similarity between RA and the broader primary motor cortex in humans is due in part to similarity with its layer 5 neurons, and is separate from the similarities to LMC/dLSC within primary motor cortex.

In situ hybridization of 15 of the 55 genes (selected because of their axon guidance functions or previously published profiles) revealed a ~95% true positive rate for differential expression in zebra finch RA relative to one or more parts of the adjacent arcopallium (Fig. 6, figs. S9 and S10, and table S5). The majority of identified genes (50 genes; ~91%) had decreased expression in the RA analogs and human LMC/dLSC relative to the surrounding regions, as opposed to only four that had increased levels (Fig. 6). This is in contrast to Area X and the analogous putamen/caudate regions, where more than 70%

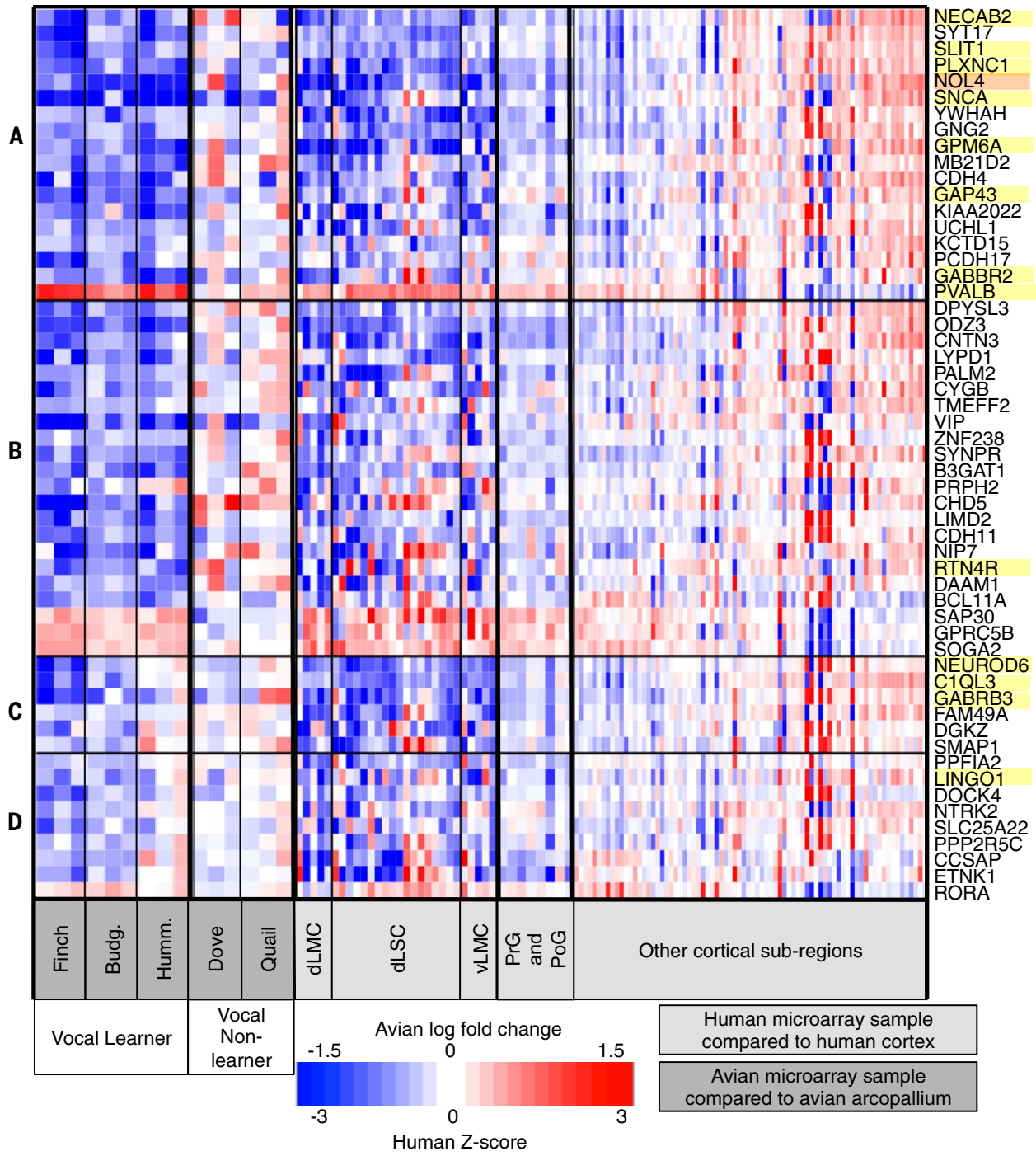


Fig. 5. Heat map of gene expression specialization in avian RA analogs versus the arcopallium and human LMC/dLSC regions versus other cortical regions. (**A** to **D**) Each row represents a gene (table S4), sorted according to whether or not they are significant in hummingbird RA analog [(A) and (B)], significant in individual human LMC/LSC regions relative to the entire PrG and PoG [(A) and (C)], or all LMC/dLSC regions (dLMC, vLMC, LSC) combined (D). Samples within each section are ordered on the basis of estimated log fold difference in LMC/LSC versus PrG and PoG surround. Each column is a microarray sample from an avian species (dark gray) or human (light gray) as listed at the bottom. For the avian species samples, each entry in the heat map shows the log fold change between each microarray sample and median gene expression value for the entire finch arcopallium (needed one species and microarray platform to normalize). For the human samples, each entry is the Z-score specialization relative to the entire cortex for the human sample or brain region. Red, higher expression; blue, lower expression; white, no difference between the compared regions. In the hummingbird microarray data set, one animal was an outlier for some genes [(C) and (D); third column], which we believe is due to an error in the laser capture dissection for a subset of sections (fig. S17); in situ hybridization data validated the hummingbird profiles of one of these genes as an example (fig. S11). Yellow highlights show validated genes; orange highlights were not able to be validated.

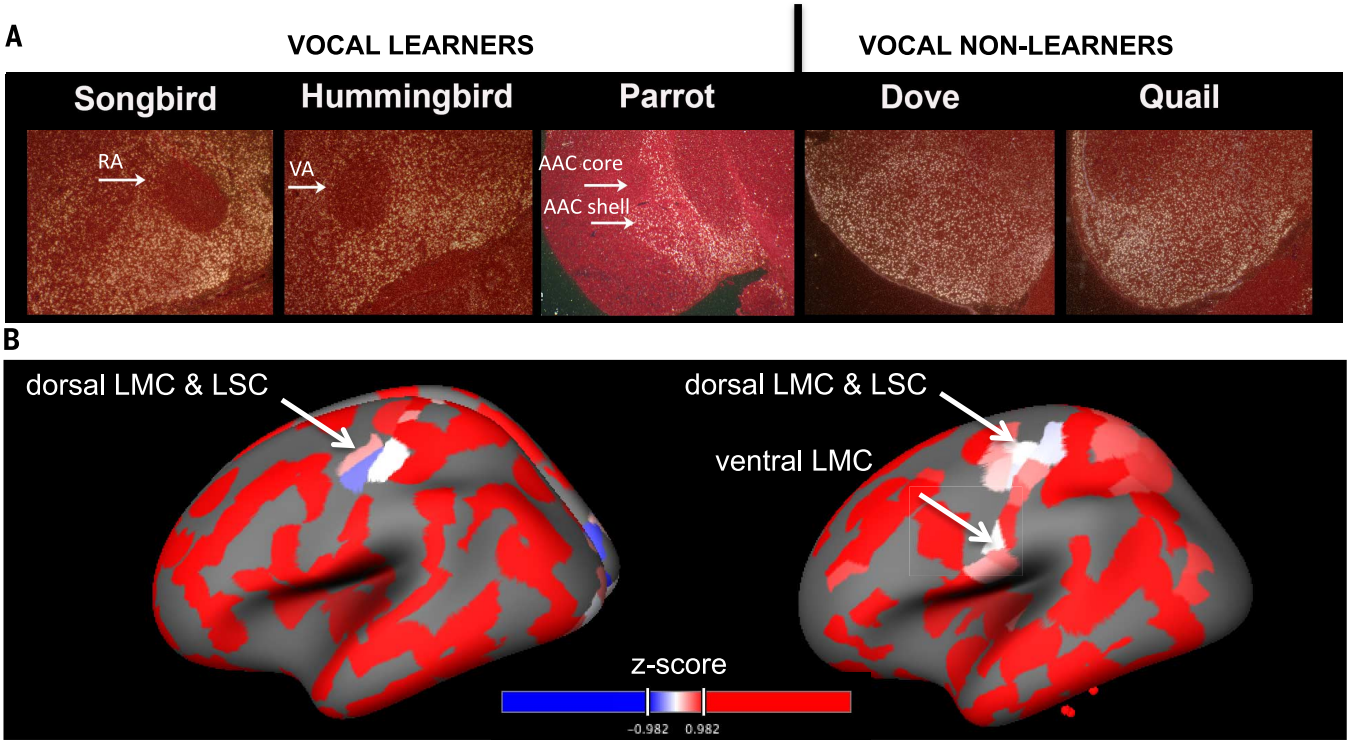


Fig. 6. Convergent differential regulation of SLIT1 in the RA analog and human LMC. (**A**) In situ hybridization of *SLIT1* in the RA analog and arcopallium of vocal-learning and vocal-nonlearning avian species. Shown are frontal sections; dorsal is up, right is midline. White, *SLIT1* mRNA detected by ^{35}S riboprobe in dark-field view; red, cresyl violet stain of brain cells. (**B**) Cortical surface map of *SLIT1* relative gene expression levels (Z-score) in the human brain measured by microarrays (http://human.brain-map.org/static/brainexplorer). Red, higher expression; blue, lower expression. Two example persons are shown (left hemisphere), one with both dLMC/LSC and vLMC. Dorsal is up, front is left.

of the genes had increased expression (fig. S7). The top enrichment categories among the 40 genes shared with motor cortex cells were movement disorders and abnormal morphology of neurons (table S5D), whereas the 55 genes shared more specifically with LMC/dLSC were related to cognitive behaviors (cognition, learning, and memory) and development of neural connections (neuritogenesis, axon and dendrite growth) for different gene subsets (5 to 15 genes each; table S5E). The latter finding suggests that the specialized genes could play a role in mediating the differential connectivity of avian RA and of human LMC.

In situ patterns of commonly specialized avian RA and human LMC genes

To achieve even better resolution, we analyzed all 55 candidate human LMC/dLSC genes with the use of microarray expression cortical surface profiles from the Allen Human Brain Atlas (http://human.brain-map.org/static/brainexplorer) and analyzed the 15 candidate avian genes by means of in situ hybridization on zebra finch brains (SM12). The anatomical profiles of the genes fell into three categories (Fig. 6 and figs. S9 and S10): (i) those with expression specialization in the RA analog and human dLMC/dLSC cortex; (ii) those with expression specialization in the RA plus immediately adjacent intermediate arcopallium and LMC plus the adjacent human primary motor cortex; and (iii) those with a more varied pattern of expression within the avian or human brain, but with local differences in avian RA and human motor cortex. On the basis of these results, we analyzed a subset of 10 of the 15 genes by in situ hybridization in our predicted dLMC/dLSC region of the other hemisphere (right) of two human brain specimens used for the microarrays; although we noted a possible stronger relationship of RA with the speech-dominant left LMC/dLSC hemisphere, the right hemisphere was also correlated (SM8).

A candidate with one of the most distinct specializations was the axon guidance ligand SLIT1, which showed verified down-regulation in the RA analog of all vocal-learning bird species, but no down-regulation in the arcopallium of the vocal-nonlearning species (Fig. 6, A and B); it was also partially down-regulated in the adjacent mAC of zebra finches. Likewise, macaques have recently been shown to have lower *SLIT1* in primary motor cortex (*69*), which we also found in marmosets (fig. S13). However, like songbird RA, our putative human dLMC in the central sulcus showed further down-regulation, particularly in the upper cortical layers. The adjacent dLSC showed striking down-regulation throughout the region (Fig. 7). In nonhuman primates, a ventral premotor region (area 6v) is hypothesized to be a precursor of the human primary LMC and makes an indirect connection to vocal motor neurons in the brainstem, However, unlike vocal-learning birds and humans, this region is not required for vocalization (*61*) and did not show differential expression of *SLIT1* (fig. S13). In parrots, the RA analog has a core portion that makes a direct projection to brainstem vocal motor neurons, like songbird and hummingbird RA analogs and the human LMC, as well as a more ventral shell portion unique to parrots that projects to other song nuclei (*70*). Only the core portion of the parrot RA analog showed *SLIT1* down-regulation (Fig. 6A) (*70*). *SLIT1* promoter is a target of the *FOXP2* transcription factor (*71*), and *FOXP2* showed lower expression in the dLMC region (fig. S12).

Another candidate gene was the *NEUROD6* transcription factor, involved in brain development and differentiation. Although weakly down-regulated in zebra finch RA, the pattern of down-regulation in the putative human dLMC was distinct such that it defined the boundaries of the region (Fig. 7). We examined negative control genes, some belonging to the same gene families (such as *SLIT2*), and did not find differential expression in RA analogs (*70*) and human LMC (Fig. 7). We also examined RNA expression patterns for 10 of the candidate genes available in the Allen Mouse Brain Atlas (http://mouse.brain-map.org) in a putative rudimentary LMC region of the mouse brain that makes a very sparse direct projection to vocal motor neurons (*72*), and did not identify examples of differential expression (fig. S14), except for *PVALB* but with reduced

expression (fig. S14); *FOXP2* was also in the opposite direction, with up-regulation in layer 5 cells of primary motor cortex (M1; fig. S14) (*73*). A comparison of expression levels in the microarray data reveals that some of these genes have some of the largest log fold expression differences among all 7473 orthologs in songbird RA and human LMC regions, with *SLIT1*, *NEUROD6*, *C1QL3*, and *PVALB* being among the most differentially expressed genes in the dLMC and RA of both species (fig. S15). A further characterization of the expression profiles of these and other genes can be found in the legends of figs. S9 to S14.

Proteomics supports difference between zebra finch RA and surrounding arcopallium

To determine whether the gene expression specializations we found in RA were reflected in the corresponding protein products, we performed nanoscale capillary tandem mass spectrometry experiments on zebra finch RA and the adjacent mAC with a microproteomic method developed for laser capture–microdissected cells (*74*). We identified 81 proteins as differentially expressed in RA (fig. S16A) and detected corresponding mRNA gene expression differences in the microarray experiment (table S7). Three of the specialized down-regulated proteins, *GAP43*, *SNCA*, and *UCHL1*, were also identified in the mRNA analyses for avian RA and human LMC. Overall, individual mRNA and protein level differences in RA and the surrounding arcopallium had a modest but significant correlation (R^2 = 0.18, $P = 9.4 \times 10^{-5}$). Notably, the genes that had lower mRNA expression in RA showed a stronger correlation with protein levels (*t* test, $P = 0.006$; fig. S16B). Concordant with the mRNA expression analysis, the top disease categories for the protein analyses included movement disorders ($P = 8.7 \times 10^{-14}$, 29 genes), and other categories included neural connectivity and brain development (table S5F). The top biological pathway gene ontology category overall was endocytosis of vesicles involved in cellular assembly and organization ($P = 6.6 \times 10^{-11}$, 7 genes), consistent with a role in neural connectivity.

Discussion

Our study indicates that behavioral and neuroanatomical convergence for the vocal-learning trait is associated with molecular convergence of gene expression changes in the circuits that control the behavior. The findings identify specific molecularly analogous brain regions for song and speech between birds and humans, and further support broader homologous brain regions in which these specialized song and speech regions are located. Although the correlations we find at first glance may not seem strong, they are highly significant, are predominantly driven by the specialized expression of tens to hundreds of genes (depending on the level in the anatomical tree) out of thousands of genes, and for vocal-learning regions are supported by lack of such findings in vocal-nonlearning birds and primates. We believe that the success of identifying these genes arises from a combination of the number of tissue sources from both birds and primates, the selection of brain regions we profiled, and the computational approach we developed, based on an anatomical gene expression specialization tree.

We find the convergent gene expression for RA/LMC and Area X/putamen remarkable in that they involve species separated from a common ancestor by 68 million to 310 million years (*34*–*36*), different vertebrate lineages, and many genes. Consistent with the function of the genes in neural connectivity and motor behavior, the avian RA and human primary motor cortex make more direct contacts with vocal motor neurons than do motor cortex neurons in any other vertebrates tested to date (*6*, *61*). Direct cortical projections to lower motor neurons are correlated with finer motor control and more complex motor learning (*75*). In contrast, the avian arcopallium, like the motor cortex in many mammals, is thought to make mostly indirect contact with lower motor neurons (*6*). We caution that our analyses do not show that the specializations we discovered are functionally related to the projections of these neurons or to vocal learning, because the difference in gene expression could be a consequence, rather than a cause, of the differential connectivity within and across species. Nonetheless, even if only a consequence, the genes and their expression differences can serve as candidates that can lead us to the cause.

Our findings on *SLIT1* add to a growing body of studies implicating an important role for this gene in speech and vocal learning. *SLIT1* was found to be more differentially up-regulated by the human *FOXP2* transcription factor than by the chimpanzee *FOXP2* (*71*). SLIT1 is a ligand for the ROBO1 axon guidance receptor, and *ROBO1* mutations cause dyslexia and speech disorders in humans (*76*). *ROBO1* is one of five candidate genes with convergent amino acid substitutions in vocal-learning mammals (*32*). *ROBO1* itself was also found to be differentially expressed in fetal human speech brain regions (*77*) and in the RA analog of avian vocal learners (*70*). Thus, because binding of SLIT1 to ROBO1 causes repulsion of the axonal processes from cell bodies (*78*), we hypothesize that the down-regulation of *SLIT1* and other neural connectivity genes in RA and LMC may represent a permissive mechanism to allow certain neural connections to form, such as the unique direct projection to vocal motor neurons.

Among the very few genes with convergent up-regulation in RA and LMC, the calcium-buffering protein parvalbumin has been noted (*16*) to be present at unusually high levels in the RA neurons that project to vocal motor neurons in songbirds and in the analogous layer 5 neurons of oral motor cortex (containing vLMC) of humans and chimpanzees, but not in the projection neurons of the adjacent arcopallium or in the oral motor cortex of other primates or mammals (*79*, *80*). Normally, parvalbumin is expressed in fast-spiking inhibitory interneurons of the cortex. It has been hypothesized that the increased expression in RA projection neurons (*16*), and we propose here also in LMC, is to protect these brain regions from Ca^{2+} toxicity due to highly active neurons that control some of the fastest muscles in the body (syrinx in birds and larynx in humans). This is supported by the fact that the RA neurons have a higher spontaneous firing rate than the surrounding arcopallium (*81*).

The enriched neurotransmission functions of the convergently specialized genes in songbird Area X and parts of the human putamen/caudate are consistent with findings that Area X does not appear to have large differences in neural connectivity relative to the surrounding striatum (*82*), but like in RA, the neural activity in Area X is commonly known to differ. The match of Area X to the human striatum and of RA analogs to the cortex enables specific hypotheses about which brain regions are analogous in song-learning birds and humans, as well as which regions in humans may be specialized for speech. The role of the human striatum in speech acquisition and production has begun to be appreciated (*57*, *83*), but specific loci have not been as well defined as for the cortex. Similarly, the dLMC was also only recently described (*63*), and our results add to the growing body of work (*62*) that shows this region could viably be a specialized part of the speech control circuit.

Our lack of finding significant convergent specializations of the songbird HVC or LMAN with humans does not mean that shared specializations of more specific regions within the cortex do not exist. Identifying these may require analyses of the transcriptomes of the adjacent non-song neurons in the avian nidopallium in which HVC and LMAN are located; for RA and Area X, in contrast, the immediate adjacent areas were profiled. We attempted to analyze an available data set of HVC and the surrounding auditory nidopallium shelf, generated with cDNA microarrays (*12*). However, the overlap between our Agilent microarray was not high enough to incorporate these data into our framework.

At higher levels of brain organization, our cortical results are in contrast with findings of a recent study that used brain transcriptome profiles to argue that the avian pallium is equally and only weakly similar to the mouse cortex, claustrum, or amygdala (*68*). We believe our ability to identify a stronger relationship to the cortex was because we organized the expression profiles as specialization trees, developed a dynamic programming algorithm requiring hierarchical anatomical alignment of brain similarities, and performed avian brain dissections using more precise laser capture microdissections. It is also possible that even though the pallial song nuclei (which made up 75% of the avian pallial samples) have more convergent similarities with the entire human cortex, the avian pallial subdivisions in which the song nuclei are housed have similarities with the claustrum and amygdala. However, our results across species suggest that

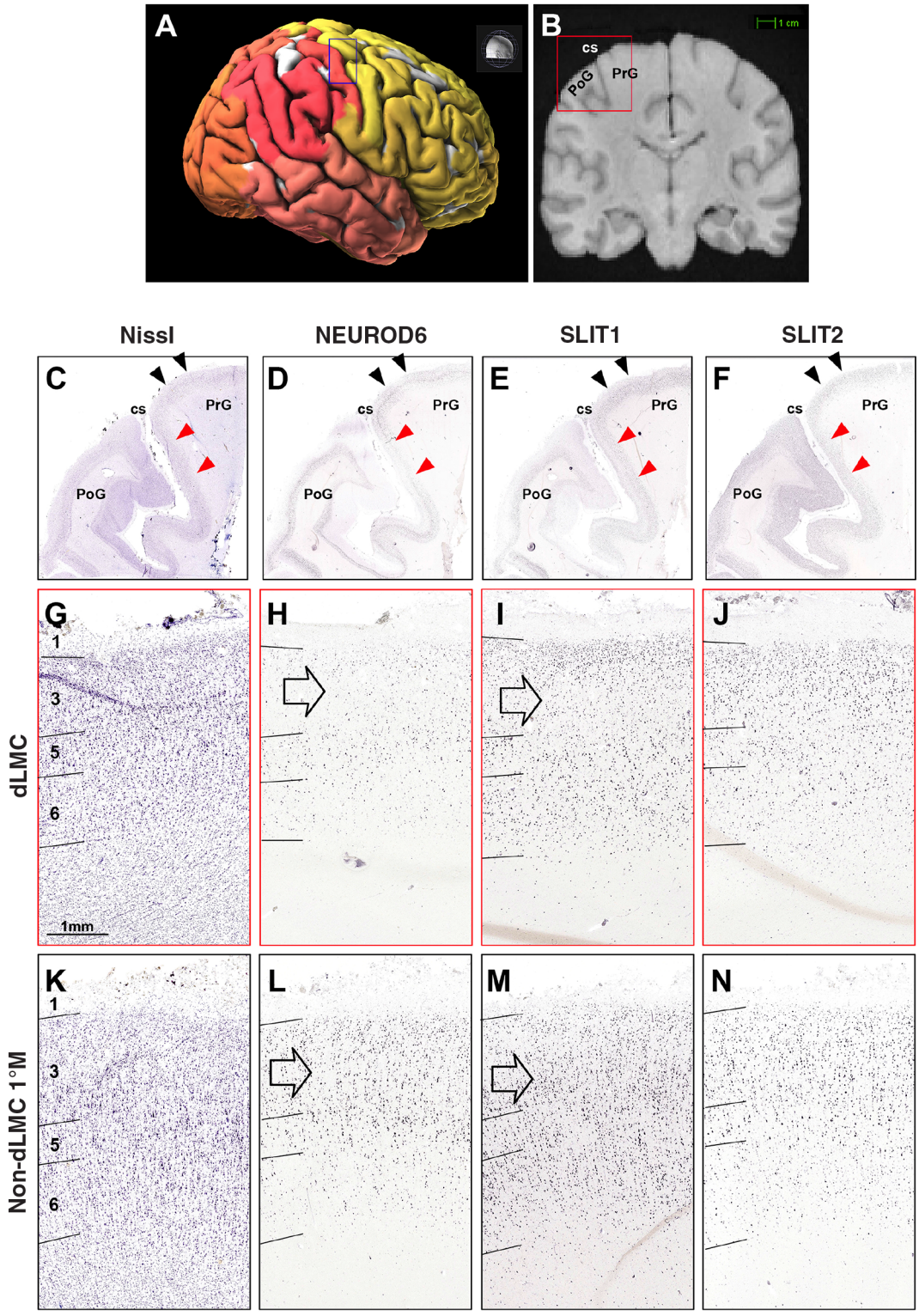


Fig. 7. In situ hybridization localization of the putative dLMC/dLSC in the human brain. (**A**) Surface image of a human brain showing the different cortical lobes (colors) and the region dissected for the in situ hybridization analyses (box). (**B**) Magnetic resonance image showing the location of the region dissected for in situ hybridization analyses in the right hemisphere (box). (**C**) Nissl stain of the examined region. (**D**) *NEUROD6* down-regulation in a distinct region of the PrG, and in the upper layers of the adjacent PoG. (**E**) *SLIT1* down-regulation in the same PrG dLMC region (see I versus M), as well as in the adjacent PoG. (**F**) *SLIT2* control showing no noticeable difference. (**G** to **N**) Red arrows correspond to the boundaries of the regions represented in the higher-power images of (G) to (J); black arrows correspond to (K) to (N). Down-regulation in dLMC is strongest in layer 3 (open arrow), but also in layer 5 for *NEUROD6*.

this is not the case for the arcopallium, the region most often suggested to be similar to the mammalian amygdala (*34*, *44*).

Our study has identified molecularly analogous brain regions that function in song and speech in vocal-learning birds and humans. The gene expression similarities we find across species, combined with circuit and functional similarities (*40*, *41*), suggest that avian vocal learners can be used as models for speech production at a molecular level. Our experiments provide a candidate set of genes involved in neural connectivity and cell communication functions. Future studies will be necessary to decipher the specific function of these genes in vocal communication and the mechanisms through which they are differentially expressed and selected. Insight into the latter question is beginning to be revealed, as in our companion study we found convergent mutations in the genomic regulatory regions of some of the genes we find here with specialized expression in the RA analog across vocal-learning lineages (*33*). Our computational approach can also be applied to other

anatomical molecular profiling experiments across species. The fact that convergent neural circuits for vocal learning are accompanied by convergent molecular changes of multiple genes in species separated by millions of years from a common ancestor indicates that brain circuits for complex traits may have limited ways in which they could have evolved from a common ancestor.

REFERENCES AND NOTES

1. C. I. Petkov, E. D. Jarvis, Birds, primates, and spoken language origins: Behavioral phenotypes and neurobiological substrates. *Front. Evol. Neurosci.* **4**, 12 (2012). doi: 10.3389/fnevo.2012.00012; pmid: 22912615
2. U. Jürgens, The neural control of vocalization in mammals: A review. *J. Voice* **23**, 1–10 (2009). doi: 10.1016/j.jvoice.2007.07.005; pmid: 18207362
3. S. E. Egnor, M. D. Hauser, A paradox in the evolution of primate vocal learning. *Trends Neurosci.* **27**, 649–654 (2004). doi: 10.1016/j.tins.2004.08.009; pmid: 15474164
4. A. J. Doupe, P. K. Kuhl, Birdsong and human speech: Common themes and mechanisms. *Annu. Rev. Neurosci.* **22**, 567–631 (1999). doi: 10.1146/annurev.neuro.22.1.567; pmid: 10202549
5. J. J. Bolhuis, K. Okanoya, C. Scharff, Twitter evolution: Converging mechanisms in birdsong and human speech. *Nat. Rev. Neurosci.* **11**, 747–759 (2010). doi: 10.1038/nrn2931; pmid: 20959859
6. G. Arriaga, E. D. Jarvis, Mouse vocal communication system: Are ultrasounds learned or innate? *Brain Lang.* **124**, 96–116 (2013). doi: 10.1016/j.bandl.2012.10.002; pmid: 23295209
7. T. H. Oakley, A. S. Rivera, Genomics and the evolutionary origins of nervous system complexity. *Curr. Opin. Genet. Dev.* **18**, 479–492 (2008). doi: 10.1016/j.gde.2008.12.002; pmid: 19152785
8. H. C. Barrett, A hierarchical model of the evolution of human brain specializations. *Proc. Natl. Acad. Sci. U.S.A.* **109** (suppl. 1), 10733–10740 (2012). doi: 10.1073/pnas.1201898109; pmid: 22723350
9. G. Feenders *et al*., Molecular mapping of movement-associated areas in the avian brain: A motor theory for vocal learning origin. *PLOS ONE* **3**, e1768 (2008). doi: 10.1371/journal.pone.0001768
10. D. Margoliash *et al*., Distributed representation in the song system of oscines: Evolutionary implications and functional consequences. *Brain Behav. Evol.* **44**, 247–264 (1994). doi: 10.1159/000113580; pmid: 7842284
11. C. V. Mello, G. E. Vates, S. Okuhata, F. Nottebohm, Descending auditory pathways in the adult male zebra finch (*Taeniopygia guttata*). *J. Comp. Neurol.* **395**, 137–160 (1998). doi: 10.1002/(SICI)1096-9861(19980601)395:2<137::AID-CNE1>3.0.CO;2-3; pmid: 9603369
12. P. V. Lovell, D. F. Clayton, K. L. Replogle, C. V. Mello, Birdsong "transcriptomics": Neurochemical specializations of the oscine song system. *PLOS ONE* **3**, e3440 (2008). doi: 10.1371/journal.pone.0003440; pmid: 18941504
13. P. V. Lovell, J. B. Carleton, C. V. Mello, Genomics analysis of potassium channel genes in songbirds reveals molecular specializations of brain circuits for the maintenance and production of learned vocalizations. *BMC Genomics* **14**, 470 (2013). doi: 10.1186/1471-2164-14-470; pmid: 23845108
14. E. Matsunaga, K. Suzuki, T. Kobayashi, K. Okanoya, Comparative analysis of mineralocorticoid receptor expression among vocal learners (Bengalese finch and budgerigar) and non-vocal learners (quail and ring dove) has implications for the evolution of avian vocal learning. *Dev. Growth Differ.* **53**, 961–970 (2011). doi: 10.1111/j.1440-169X.2011.01302.x; pmid: 22010640
15. A. T. Hilliard, J. E. Miller, S. Horvath, S. A. White, Distinct neurogenomic states in basal ganglia subregions relate differently to singing behavior in songbirds. *PLOS Comput. Biol.* **8**, e1002773 (2012). doi: 10.1371/journal.pcbi.1002773; pmid: 23144607
16. E. Hara, M. V. Rivas, J. M. Ward, K. Okanoya, E. D. Jarvis, Convergent differential regulation of parvalbumin in the brains of vocal learners. *PLOS ONE* **7**, e29457 (2012). doi: 10.1371/journal.pone.0029457; pmid: 22238614
17. H. Horita *et al*., Specialized motor-driven dusp1 expression in the song systems of multiple lineages of vocal learning birds. *PLOS ONE* **7**, e42173 (2012). doi: 10.1371/journal.pone.0042173; pmid: 22876306
18. K. Amunts *et al*., Broca's region: Novel organizational principles and multiple receptor mapping. *PLOS Biol.* **8**, e1000489 (2010). doi: 10.1371/journal.pbio.1000489; pmid: 20877713
19. K. Wada, H. Sakaguchi, E. D. Jarvis, M. Hagiwara, Differential expression of glutamate receptors in avian neural pathways for learned vocalization. *J. Comp. Neurol.* **476**, 44–64 (2004). doi: 10.1002/cne.20201; pmid: 15236466
20. C. S. Lai, S. E. Fisher, J. A. Hurst, F. Vargha-Khadem, A. P. Monaco, A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* **413**, 519–523 (2001). doi: 10.1038/35097076; pmid: 11586359
21. S. Haesler *et al*., FoxP2 expression in avian vocal learners and non-learners. *J. Neurosci.* **24**, 3164–3175 (2004). doi: 10.1523/JNEUROSCI.4369-03.2004; pmid: 15056696
22. S. Haesler *et al*., Incomplete and inaccurate vocal imitation after knockdown of FoxP2 in songbird basal ganglia nucleus Area X. *PLOS Biol.* **5**, e321 (2007). doi: 10.1371/journal.pbio.0050321; pmid: 18052609
23. C. Scharff, J. Petri, Evo-devo, deep homology and FoxP2: Implications for the evolution of speech and language. *Philos. Trans. R. Soc. London Ser. B* **366**, 2124–2140 (2011). doi: 10.1098/rstb.2011.0001; pmid: 21690130
24. I. Teramitsu, L. C. Kudo, S. E. London, D. H. Geschwind, S. A. White, Parallel FoxP1 and FoxP2 expression in songbird and human brain predicts functional interaction. *J. Neurosci.* **24**, 3152–3163 (2004). doi: 10.1523/JNEUROSCI.5589-03.2004; pmid: 15056695
25. S. Yokoyama, F. B. Radlwimmer, The molecular genetics and evolution of red and green color vision in vertebrates. *Genetics* **158**, 1697–1710 (2001). pmid: 11545071
26. J. Arendt, D. Reznick, Convergence and parallelism reconsidered: What have we learned about the genetics of adaptation? *Trends Ecol. Evol.* **23**, 26–32 (2008). doi: 10.1016/j.tree.2007.09.011; pmid: 18022278
27. P. J. Wittkopp, B. L. Williams, J. E. Selegue, S. B. Carroll, *Drosophila* pigmentation evolution: Divergent genotypes underlying convergent phenotypes. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 1808–1813 (2003). doi: 10.1073/pnas.0336368100; pmid: 12574518
28. M. D. Shapiro *et al*., Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* **428**, 717–723 (2004). doi: 10.1038/nature02415; pmid: 15085123
29. H. H. Zakon, Y. Lu, D. J. Zwickl, D. M. Hillis, Sodium channel genes and the evolution of diversity in communication signals of electric fishes: Convergent molecular evolution. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 3675–3680 (2006). doi: 10.1073/pnas.0600160103; pmid: 16505358
30. Y. Liu *et al*., Convergent sequence evolution between echolocating bats and dolphins. *Curr. Biol.* **20**, R53–R54 (2010). doi: 10.1016/j.cub.2009.11.058; pmid: 20129036
31. J. Parker *et al*., Genome-wide signatures of convergent evolution in echolocating mammals. *Nature* **502**, 228–231 (2013). doi: 10.1038/nature12511; pmid: 24005325
32. R. Wang, thesis, Duke University (2011); http://dukespace.lib.duke.edu/dspace/bitstream/handle/10161/5630/Wang_duke_0066D_10911.pdf?sequence=1.
33. G. Zhang *et al*., Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* **346**, 1311–1320 (2014). doi: 10.1126/science.1251385
34. E. D. Jarvis *et al*., Avian brains and a new understanding of vertebrate brain evolution. *Nat. Rev. Neurosci.* **6**, 151–159 (2005). doi: 10.1038/nrn1606; pmid: 15685220
35. S. Kumar, S. B. Hedges, A molecular timescale for vertebrate evolution. *Nature* **392**, 917–920 (1998). doi: 10.1038/31927; pmid: 9582070
36. E. D. Jarvis *et al*., Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* **346**, 1320–1331 (2014). doi: 10.1126/science.1253451
37. O. Whitney *et al*., *Science* **346**, 1256780 (2014). doi: 10.1126/science.1256780
38. M. J. Hawrylycz *et al*., An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* **489**, 391–399 (2012). doi: 10.1038/nature11405; pmid: 22996553
39. A. Bernard *et al*., Transcriptional architecture of the primate neocortex. *Neuron* **73**, 1083–1099 (2012). doi: 10.1016/j.neuron.2012.03.002; pmid: 22445337
40. M. S. Brainard, A. J. Doupe, Translating birdsong: Songbirds as a model for basic and applied medical research. *Annu. Rev. Neurosci.* **36**, 489–517 (2013). doi: 10.1146/annurev-neuro-060909-152826; pmid: 23750515
41. E. D. Jarvis, Learned birdsong and the neurobiology of human language. *Ann. N.Y. Acad. Sci.* **1016**, 749–777 (2004). doi: 10.1196/annals.1298.038; pmid: 15313804
42. E. D. Jarvis, F. Nottebohm, Motor-driven gene expression. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 4097–4102 (1997). doi: 10.1073/pnas.94.8.4097; pmid: 9108111
43. E. D. Jarvis *et al*., Global view of the functional molecular organization of the avian cerebrum: Mirror images and functional columns. *J. Comp. Neurol.* **521**, 3614–3665 (2013). pmid: 23818122
44. A. Reiner *et al*., The Avian Brain Nomenclature Forum: Terminology for a new century in comparative neuroanatomy. *J. Comp. Neurol.* **473**, E1–E6 (2004). pmid: 19626136
45. L. Puelles *et al*., Pallial and subpallial derivatives in the embryonic chick and mouse telencephalon, traced by the expression of the genes Dlx-2, Emx-1, Nkx-2.1, Pax-6, and Tbr-1. *J. Comp. Neurol.* **424**, 409–438 (2000). doi: 10.1002/1096-9861(20000828)424:3<409::AID-CNE3>3.0.CO;2-7; pmid: 10906711
46. G. D. Carrillo, A. J. Doupe, Is the songbird Area X striatal, pallidal, or both? An anatomical study. *J. Comp. Neurol.* **473**, 415–437 (2004). doi: 10.1002/cne.20099; pmid: 15116398
47. A. Reiner, A. V. Laverghetta, C. A. Meade, S. L. Cuthbertson, S. W. Bottjer, An immunohistochemical and pathway tracing study of the striatopallidal organization of area X in the male zebra finch. *J. Comp. Neurol.* **469**, 239–261 (2004). doi: 10.1002/cne.11012; pmid: 14694537
48. M. A. Farries, D. J. Perkel, A telencephalic nucleus essential for song learning contains neurons with physiological characteristics of both striatum and globus pallidus. *J. Neurosci.* **22**, 3776–3787 (2002). pmid: 11978853
49. F. Nottebohm, T. M. Stokes, C. M. Leonard, Central control of song in the canary, Serinus canarius. *J. Comp. Neurol.* **165**, 457–486 (1976). doi: 10.1002/cne.901650405; pmid: 1262540
50. J. A. Rathelot, P. L. Strick, Subdivisions of primary motor cortex based on cortico-motoneuronal cells. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 918–923 (2009). doi: 10.1073/pnas.0808362106; pmid: 19139417
51. Y. Mandelblat-Cerf, L. Las, N. Denisenko, M. S. Fee, A role for descending auditory cortical projections in songbird vocal learning. *eLife* **3**, e04371 (2014). doi: 10.7554/eLife.04371; pmid: 24935934
52. J. L. Dubbeldam, The neural substrate for 'learned' and 'nonlearned' activities in birds: A discussion of the organization of bulbar reticular premotor systems with side-lights on the mammalian situation. *Acta Anat.* **163**, 157–172 (1998). doi: 10.1159/000046494; pmid: 9973636
53. S. W. Bottjer, J. D. Brady, B. Cribbs, Connections of a motor cortical region in zebra finches: Relation to pathways for vocal learning. *J. Comp. Neurol.* **420**, 244–260 (2000). doi: 10.1002/(SICI)1096-9861(20000501)420:2<244::AID-CNE7>3.0.CO;2-M; pmid: 10753310
54. J. M. Wild, D. Li, C. Eagleton, Projections of the dorsomedial nucleus of the intercollicular complex (DM) in relation to respiratory-vocal nuclei in the brainstem of pigeon (*Columba livia*) and zebra finch (*Taeniopygia guttata*). *J. Comp. Neurol.* **377**, 392–413 (1997). doi: 10.1002/(SICI)1096-9861(19970120)377:3<392::AID-CNE7>3.0.CO;2-Y; pmid: 8989654
55. Y. Nii, S. Uematsu, R. P. Lesser, B. Gordon, Does the central sulcus divide motor and sensory functions? Cortical mapping of human hand areas as revealed by electrical stimulation through subdural grid electrodes. *Neurology* **46**, 360–367 (1996). doi: 10.1212/WNL.46.2.360; pmid: 8614495
56. P. Sörös *et al*., Clustered functional MRI of overt speech production. *Neuroimage* **32**, 376–387 (2006). doi: 10.1016/j.neuroimage.2006.02.046; pmid: 16631384
57. S. Brown *et al*., The somatotopy of speech: Phonation and articulation in the human motor cortex. *Brain Cognit.* **70**, 31–41 (2009). doi: 10.1016/j.bandc.2008.12.006; pmid: 19162389
58. J. W. Bohland, F. H. Guenther, An fMRI investigation of syllable sequence production. *Neuroimage* **32**, 821–841 (2006). doi: 10.1016/j.neuroimage.2006.04.173; pmid: 16730195
59. K. Grabski *et al*., Functional MRI assessment of orofacial articulators: Neural correlates of lip, jaw, larynx, and tongue movements. *Hum. Brain Mapp.* **33**, 2306–2321 (2012). doi: 10.1002/hbm.21363; pmid: 21826760
60. K. Simonyan, J. Ostuni, C. L. Ludlow, B. Horwitz, Functional but not structural networks of the human laryngeal motor cortex show left hemispheric lateralization during syllable but not breathing production. *J. Neurosci.* **29**, 14912–14923 (2009). doi: 10.1523/JNEUROSCI.4897-09.2009; pmid: 19940187
61. K. Simonyan, B. Horwitz, Laryngeal motor cortex and control of speech in humans. *Neuroscientist* **17**, 197–208 (2011). doi: 10.1177/1073858410386727; pmid: 21362688

62. K. Simonyan, The laryngeal motor cortex: Its organization and connectivity. *Curr. Opin. Neurobiol.* **28**, 15–21 (2014). doi: 10.1016/j.conb.2014.05.006; pmid: 24929930
63. S. Brown, E. Ngan, M. Liotti, A larynx area in the human motor cortex. *Cereb. Cortex* **18**, 837–845 (2008). doi: 10.1093/cercor/bhm131; pmid: 17652461
64. F. Nottebohm, M. E. Nottebohm, Vocalizations and breeding behaviour of surgically deafened ring doves (*Streptopelia risoria*). *Anim. Behav.* **19**, 313–327 (1971). doi: 10.1016/S0003-3472(71)80012-X; pmid: 5150477
65. S. Derégnaucourt, S. Saar, M. Gahr, Dynamics of crowing development in the domestic Japanese quail (*Coturnix coturnix japonica*). *Proc. Biol. Sci.* **276**, 2153–2162 (2009). doi: 10.1098/rspb.2009.0016; pmid: 19324760
66. E. D. Jarvis *et al*., A framework for integrating the songbird brain. *J. Comp. Physiol. A* **188**, 961–980 (2002). doi: 10.1007/s00359-002-0358-y; pmid: 12471494
67. H. J. Karten, Homology and evolutionary origins of the 'neocortex'. *Brain Behav. Evol.* **38**, 264–272 (1991). doi: 10.1159/000114393; pmid: 1777808
68. T. G. Belgard *et al*., Adult pallium transcriptomes surprise in not reflecting predicted homologies across diverse chicken and mouse pallial sectors. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 13150–13155 (2013). doi: 10.1073/pnas.1307444110; pmid: 23878249
69. T. Sasaki, Y. Komatsu, A. Watakabe, K. Sawada, T. Yamamori, Prefrontal-enriched SLIT1 expression in Old World monkey cortex established during the postnatal development. *Cereb. Cortex* **20**, 2496–2510 (2010). doi: 10.1093/cercor/bhp319; pmid: 20123755
70. R. Wang *et al*., Convergent differential regulation of ROBO-SLIT axon guidance pathway genes in the brains of vocal learners. *J. Comp. Neurol.* 10.1002/cne.23719 (2014).
71. G. Konopka *et al*., Human-specific transcriptional regulation of CNS development genes by FOXP2. *Nature* **462**, 213–217 (2009). doi: 10.1038/nature08549; pmid: 19907493
72. G. Arriaga, E. P. Zhou, E. D. Jarvis, Of mice, birds, and men: The mouse ultrasonic song system has some features similar to humans and song-learning birds. *PLOS ONE* **7**, e46610 (2012). doi: 10.1371/journal.pone.0046610 pmid: 23071596
73. T. Hisaoka, Y. Nakamura, E. Senba, Y. Morikawa, The forkhead transcription factors, Foxp1 and Foxp2, identify different subpopulations of projection neurons in the mouse cerebral cortex. *Neuroscience* **166**, 551–563 (2010). doi: 10.1016/j.neuroscience.2009.12.055; pmid: 20040367
74. P. L. Roulhac, J. M. Ward, J. W. Thompson, E. J. Soderblom, M. Silva, M. A. Mosely III, E. D. Jarvis, *Cold Spring Harb. Protoc.* pdb.prot5573 (2011). doi: 10.1101/pdb.prot5573
75. R. N. Lemon, Descending pathways in motor control. *Annu. Rev. Neurosci.* **31**, 195–218 (2008). doi: 10.1146/annurev.neuro.31.060407.125547; pmid: 18558853
76. S. E. Fisher, C. Francks, Genes, cognition and dyslexia: Learning to read the genome. *Trends Cognit. Sci.* **10**, 250–257 (2006). doi: 10.1016/j.tics.2006.04.003; pmid: 16675285
77. M. B. Johnson *et al*., Functional and evolutionary insights into human brain development through global transcriptome analysis. *Neuron* **62**, 494–509 (2009). doi: 10.1016/j.neuron.2009.03.027; pmid: 19477152
78. M. E. Cornide-Petronio, A. Barreiro-Iglesias, Role of Slit and Robo proteins in the development of dopaminergic neurons. *Dev. Neurosci.* **35**, 285–292 (2013). doi: 10.1159/000351023; pmid: 23796505
79. J. M. Wild, M. N. Williams, R. A. Suthers, Parvalbumin-positive projection neurons characterise the vocal premotor pathway in male, but not female, zebra finches. *Brain Res.* **917**, 235–252 (2001). doi: 10.1016/S0006-8993(01)02938-9; pmid: 11640910
80. C. C. Sherwood, R. L. Holloway, J. M. Erwin, P. R. Hof, Cortical orofacial motor representation in Old World monkeys, great apes, and humans. II. Stereologic analysis of chemoarchitecture. *Brain Behav. Evol.* **63**, 82–106 (2004). doi: 10.1159/000075673; pmid: 14685003
81. P. Adret, D. Margoliash, Metabolic and neural activity in the song system nucleus robustus archistriatalis: Effect of age and gender. *J. Comp. Neurol.* **454**, 409–423 (2002). doi: 10.1002/cne.10459; pmid: 12455006
82. A. J. Doupe, D. J. Perkel, A. Reiner, E. A. Stern, Birdbrains could teach basal ganglia research a new song. *Trends Neurosci.* **28**, 353–363 (2005). doi: 10.1016/j.tins.2005.05.005; pmid: 15935486
83. K. Simonyan, B. Horwitz, E. D. Jarvis, Dopamine regulation of human speech and bird song: A critical review. *Brain Lang.* **122**, 142–150 (2012). doi: 10.1016/j.bandl.2011.12.009; pmid: 22284300
84. K. Okanoya, B. Merker, in *Emergence of Communication and Language*, C. Lyon, C. L. Nehaniv, A. Cangelosi, Eds. (Springer, London, 2007), pp. 421–434.
85. H. J. Karten, T. Shimizu, The origins of neocortex: Connections and lamination as distinct events in evolution. *J. Cognit. Neurosci.* **1**, 291–301 (1989). doi: 10.1162/jocn.1989.1.4.291; pmid: 23971981

ACKNOWLEDGMENTS

We thank H. Dressman and L.-L. Rowlette for microarray hybridizations, the Allen Institute for Brain Science for early access to human microarray data and human brain in situ hybridizations (and the Allen Institute founders, P. G. Allen and J. Allen, for their encouragement of that project), K. Simonyan for her coordinates of meta-analyses for speech regions, and S. Jarvis for help with imaging. Supported by NIH Directors Pioneer Award DP1 OD000448, National Institute on Deafness and Other Communication Disorders grant R01DC007218, and the Howard Hughes Medical Institute (E.D.J.). The brain gene expression atlas (ZEBrA) is supported by National Institute of General Medical Sciences grant R24GM092842 (C.V.M.). Marmoset experiments were supported by the Funding Program for World-leading Innovative R&D on Science and Technology (A.I.). Microarray data are available in the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo) under accession numbers GSE33365 and GSE34819 (for songbird Area X), GSE28395 (for avian arcopallium regions), GSE33667 (for avian brainstem regions), and GSE31613 (for macaque brain regions); human microarray data are available from the Allen Human Brain Atlas site (http://human.brain-map.org/static/download). E.D.J., A.R.P., E.H., M.R., and P.R. conceived and designed the project; E.D.J., A.R.P., and A.J.H. wrote the paper; E.D.J., A.J.H., M.A.M., A.I., E.L., and C.V.M. co-supervised parts of the project; A.R.P. conducted most computational analyses; G.G., T.B., E.J.S., and A.J.H. conducted other computational analyses and supervision; E.H., M.R., O.W., and J.T.H. collected avian samples and performed microarray experiments; J.M., M.W., P.V.L., E.H., R.W., M.K., A.Bo., A.Be., C.V.M., and E.D.J. performed in situ hybridizations and analyses across species; G.Z., M.T.P.G., and E.D.J. provided genomes; and P.R., M.A.M., J.W.T., and E.J.S. performed proteomic experiments.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/346/6215/1256846/suppl/DC1
Materials and Methods
Figs. S1 to S19
Tables S1 to S8
References (*86*–*115*)

2 June 2014; accepted 11 November 2014
10.1126/science.1256846

RESEARCH ARTICLE SUMMARY

AVIAN GENOMICS

Core and region-enriched networks of behaviorally regulated genes and the singing genome

Osceola Whitney,*† Andreas R. Pfenning,*† Jason T. Howard, Charles A. Blatti, Fang Liu, James M. Ward, Rui Wang, Jean-Nicolas Audet, Manolis Kellis, Sayan Mukherjee, Saurabh Sinha, Alexander J. Hartemink, Anne E. West,* Erich D. Jarvis*

INTRODUCTION: Brain activity drives both behavior and regulated gene expression in neurons. Although past studies have identified activity-induced signaling and gene regulation cascades in cultured neurons, much less is known about how activity-dependent transcriptional networks are affected by the variations in cell-type composition, network interconnections, and firing patterns that comprise behaviorally active brain circuits in vivo.

RATIONALE: We tested the hypothesis that behaviorally regulated gene expression is anatomically and temporally diverse and that the key determinants of this diversity are networks of transcription factors, their genomic binding sites, and epigenetic chromatin states. We analyzed genome-wide, singing-regulated gene expression across time in the four major forebrain regions of the song control system in songbirds, a model of speech production in humans. We then performed a transcription factor motif analysis to identify gene regulatory networks enriched in each song nucleus and measured acetylation of histone 3 at lysine 27 (H3K27ac) to identify chromatin regions that were transcriptionally active in the genomes of song nuclei before and after singing.

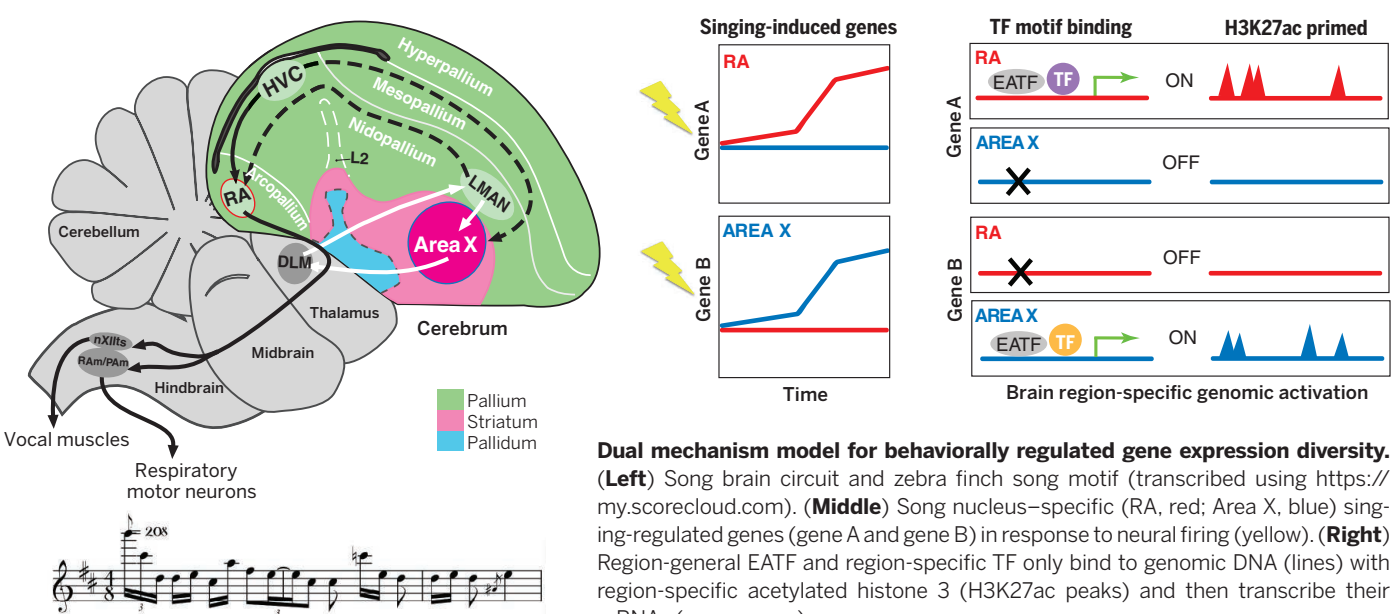
RESULTS: We found that singing was associated with differential regulation of about 10% of all genes in the avian genome that came in several waves across time. Less than 1% of these genes were comparably regulated in all song nuclei tested, and these comprised a core set dominated by immediate-early gene (IEG) transcription factors. By contrast, the vast majority of singing-regulated genes were regulated in only one or a subset of song nuclei, such that each song nucleus had its own dominant subset of genes regulated with defined temporal profiles, controlling a variety of functions. The promoters of many of the singing-regulated genes contained binding motifs for known early-activated transcription factors (EATFs) that become active in response to neural firing, some of which were expressed differentially between song nuclei at baseline. One EATF, calcium-response factor (*CaRF*), was tested with RNA interference knockdown in cultured neurons and found to regulate the predicted genes in response to neural activity, but was also found to modulate their expression even at baseline. More strikingly, we found with H3K27ac analysis that many song nucleus–specific singing-regulated genes did not show increased chromatin regulatory element activity after singing but rather already had primed region-specific regulatory activity before singing began.

ON OUR WEB SITE

Read the full article at http://dx.doi.org/10.1126/science.1256780

CONCLUSIONS: We propose a dual mechanism for the diversity of behaviorally regulated genes across different brain regions in vivo (see the figure). First, the neural activity associated with singing activates EATFs, and some TFs differentially expressed in brain regions at baseline, to trigger region-specific expression of their target genes. Second, brain region–specific enhancers near activity-regulated genes are waiting in an epigenetically primed state, ready to modulate transcription of general and song nucleus–specific genes at a moment's notice when the neurons fire. The combination of these two mechanisms underlies a great diversity of behaviorally regulated gene expression, permitting each nucleus to perform its particular function in this complex behavior. ■

The list of author affiliations is available in the full article online.
*Corresponding author. E-mail: owhitney@gmail.com (O.W.); apfenning@csail.mit.edu (A.R.P.); west@neuro.duke.edu (A.E.W.); jarvis@neuro.duke.edu (E.D.J.)
Cite this article as O. Whitney *et al.*, *Science* 346, 1256780 (2014). DOI: 10.1126/science.1256780



Dual mechanism model for behaviorally regulated gene expression diversity. (**Left**) Song brain circuit and zebra finch song motif (transcribed using https://my.scorecloud.com). (**Middle**) Song nucleus–specific (RA, red; Area X, blue) singing-regulated genes (gene A and gene B) in response to neural firing (yellow). (**Right**) Region-general EATF and region-specific TF only bind to genomic DNA (lines) with region-specific acetylated histone 3 (H3K27ac peaks) and then transcribe their mRNAs (green arrow).

PHOTO: CREDIT GOES HERE AS SHOWN; CREDIT GOES HERE AS SHOWN

Downloaded from www.sciencemag.org on December 14, 2014

Published by AAAS

RESEARCH ARTICLE

Core and region-enriched networks of behaviorally regulated genes and the singing genome

Osceola Whitney,[1]*† Andreas R. Pfenning,[1,2]*† Jason T. Howard,[1] Charles A Blatti,[3] Fang Liu,[4] James M. Ward,[1] Rui Wang,[1] Jean-Nicoles Audet,[5] Manolis Kellis,[2] Sayan Mukherjee,[6] Saurabh Sinha,[3] Alexander J. Hartemink,[7] Anne E. West,[4]* Erich D. Jarvis[1]*

Songbirds represent an important model organism for elucidating molecular mechanisms that link genes with complex behaviors, in part because they have discrete vocal learning circuits that have parallels with those that mediate human speech. We found that ~10% of the genes in the avian genome were regulated by singing, and we found a striking regional diversity of both basal and singing-induced programs in the four key song nuclei of the zebra finch, a vocal learning songbird. The region-enriched patterns were a result of distinct combinations of region-enriched transcription factors (TFs), their binding motifs, and presinging acetylation of histone 3 at lysine 27 (H3K27ac) enhancer activity in the regulatory regions of the associated genes. RNA interference manipulations validated the role of the calcium-response transcription factor (CaRF) in regulating genes preferentially expressed in specific song nuclei in response to singing. Thus, differential combinatorial binding of a small group of activity-regulated TFs and predefined epigenetic enhancer activity influences the anatomical diversity of behaviorally regulated gene networks.

Songbirds offer an important in vivo model system for studying transcriptional programs regulated during behavior. This system consists of interconnected brain nuclei that control production of a learned vocal behavior (singing) with parallels to human speech (*1*, *2*). Four key song nuclei are embedded within three regionally distinct telencephalic brain cell populations: HVC (letter based name), LMAN (lateral magnocellular nucleus of the nidopallium), RA (robust nucleus of the arcopallium), and Area X in the striatum. (Fig. 1A) (*3*–*6*). These nuclei are connected in a vocal motor pathway (HVC to RA) and a vocal learning pathway (LMAN and Area X) (*7*–*13*). Human functional analogs to these avian brain regions are in the cortex (pallium) and basal ganglia (striatum) (*2*, *6*, *14*, *15*). This includes song (avian) and speech (human) brain regions that have convergence of differentially expressed genes (*15*), which suggests that the behavioral and neuroanatomical similarities for the production of learned vocalizations are accompanied by similarities in molecular and genetic mechanisms, such as with FoxP2 (*16*).

The neural activity within song nuclei that underlies singing was initially shown to drive induction of two immediate early genes (IEGs), the transcription factors *EGR1* and *FOS* (*17*–*19*). Their levels of expression correlate with the amount of singing in a motor-driven and social-context–dependent manner (*20*–*23*). Subsequent studies identified an additional 33 genes regulated within song nuclei by singing (*24*). The identified gene products have a wide range of cellular and biological process functions (*24*), including from neurogenesis (*25*, *26*) to speech (*27*, *28*). The genes were also found to cluster in a few anatomical and short temporal patterns of expression, although this was determined manually. As a result, we hypothesized that in vivo behaviorally induced gene expression may consist of anatomically and temporally diverse gene expression programs that can be regulated by networks of combinatorial transcription factor complexes or epigenetic chromatin differences (*24*). Two reports (*29*, *30*) using our oligonucleotide microarrays found many more genes—800 to 2000 gene transcripts—regulated in the song nucleus Area X as a result of singing but could not test this hypothesis because the data were from only one song nucleus and/or one time point.

To test this hypothesis, we profiled baseline and singing-regulated gene expression across time in the four key song nuclei using our songbird gene expression microarray, which we annotated based on recently sequenced avian genomes (*15*, *31*) and the human genome. Combined with genomic transcription factor motif analyses and chromatin immunoprecipitation sequencing (ChIP-seq) detection of active chromatin, we found predominantly diverse networks of simultaneously activated cascades of behaviorally regulated genes across brain regions, which can be explained in part by a combination of transcription factor complexes and epigenetic regulatory activity in the genome.

Results

We analyzed singing-regulated gene expression at a genomic-scale in HVC, LMAN, RA, and Area X of the zebra finch (Fig. 1 and fig. S1). To do so, we recorded moment-to-moment singing behavior of all animals over a 7-hour time course, laser microdissected individual song nuclei from multiple birds at each time point, amplified their mRNA, hybridized the resulting cDNA to our custom-designed 44 K oligonucleotide microarrays (table S1), and developed a computational approach that yielded a true positive rate >87%, as verified by in situ hybridization and reverse transcription polymerase chain reaction [fig. S2; tables S2 and S3; supplementary materials sections 1 to 7 (SM1 to SM7)]. This analysis detected 24,498 expressed transcripts among the four song nuclei in silent and/or singing animals (table S4), of which 18,478 (75%) mapped to 9059 Ensembl v60 annotated genes of the zebra finch genome, indicating that at least 50% of the transcribed genome is expressed in the song-control circuit of an adult animal during awake behaving hours.

Distinct baseline gene expression profiles define the song circuit

Using a linear model that we developed to identify differentially expressed transcripts in each brain region and combinations thereof (SM6), we found that of the 24,498 transcripts, ~5167 [21%, representing 3168 genes or ~17% of the genes in the avian genome (*29*)] were differentially expressed among song nuclei at baseline in silent animals (i.e., before singing began). These 5167 transcripts were organized hierarchically into at least five major region-specific clusters (Fig. 2A and table S5) with different functional enrichments (tables S6 and S7). A striatal song nucleus (Area X) cluster was enriched with noncoding RNAs, G protein–coupled receptors, and synaptic transmission proteins (Fig. 2A, turquoise cluster, and table S6). Cortical-like song nuclei (HVC, LMAN, and RA) were enriched for cell-to-cell signaling membrane-associated, axonal connectivity, and postsynaptic density (PSD) proteins (Fig. 2A, blue cluster, and table S6). The nidopallium song nuclei (HVC and LMAN) were further enriched for another group of cell-cell communication and neural connectivity, membrane-associated proteins (Fig. 2A, yellow cluster, and table S6). The arcopallium song nucleus RA was enriched for another set of neural connectivity proteins and for proteins involved in epilepsy and Alzheimer's (Fig. 2A, green cluster, and table S6). RA was the only

[1]Department of Neurobiology, Howard Hughes Medical Institute, and Duke University Medical Center, Durham, NC 27710, USA. [2]Computer Science and Artificial Intelligence Laboratory and the Broad Institute of MIT and Harvard, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [3]Department of Computer Science, University of Illinois, Urbana-Champaign, IL, USA. [4]Department of Neurobiology, Duke University Medical Center, Durham, NC 27710, USA. [5]Department of Biology, McGill University, Montreal, Quebec H3A 1B1, Canada. [6]Department of Statistics, Duke University, Durham, NC, USA. [7]Department of Computer Science, Duke University, Durham, NC 27708-0129, USA.
*Corresponding author. E-mail: owhitney@gmail.com (O.W.); apfenning@csail.mit.edu (A.R.P.); west@neuro.duke.edu (A.E.W.); jarvis@neuro.duke.edu (E.D.J.) †These authors contributed equally to this work.

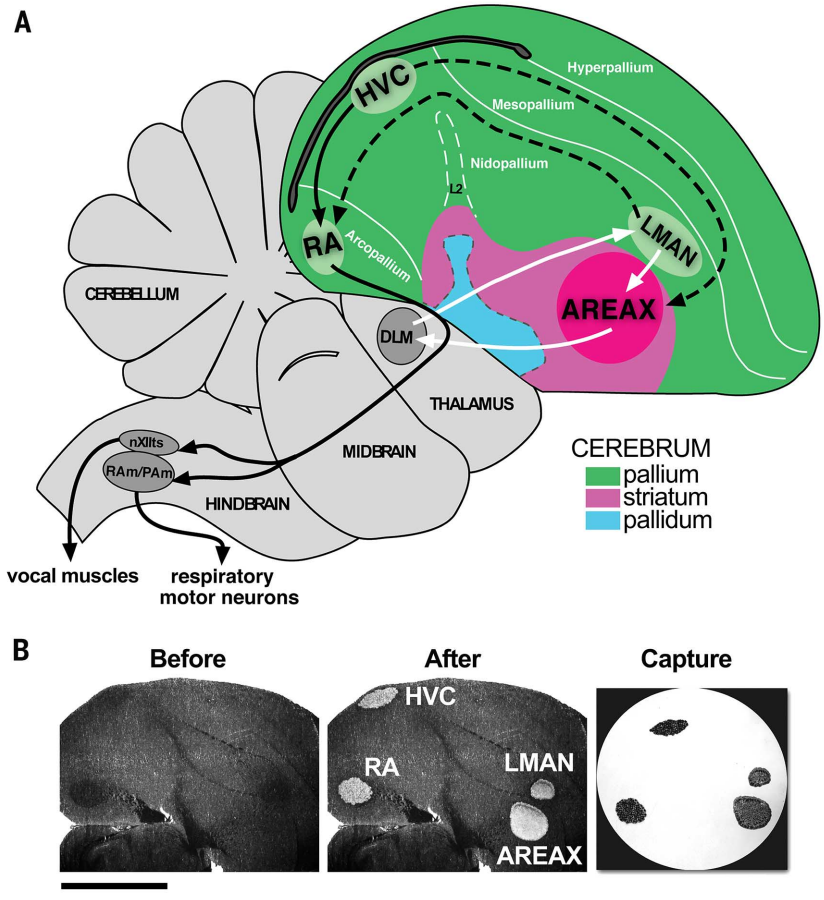


Fig. 1. Song system and laser microdissection. (**A**) Sagittal schematic of the zebra finch brain showing positions and some connections of song nuclei. Pallial, striatal, and pallidal regions are distinguished by colors. Black arrows, posterior vocal pathway involved in song production; white arrows, anterior vocal pathway involved in song learning and modulation; dashed arrows, connections between the two pathways. (**B**) Song nuclei were laser-capture microdissected from males that were either silent or continuously singing for 0.5 hours and 1 hour, and for each hour thereafter up to 7 hours, resulting in more than 200 total microarrays. Shown are images of 10-μm tissue sections before and after laser capture microdissection at 10X magnification. (Before) Following dehydration, song nuclei fiber density appears darker than surrounding tissue. (After) Song nuclei regions are selectively cut out using an infrared laser. (Capture) The cut song nuclei transferred to the cap by the LCM system. For microarray analysis, each of the four song nuclei from each animal was captured separately to individual LCM caps. Dorsal is up; anterior is right. Scale bar, 2 mm.

pallial brain region that had a large cluster of genes with a lower level of expression, which was enriched for PSD proteins different from the cortical enrichment (Fig. 2A, brown cluster, and table S6), and LMAN was the only song nucleus that did not have a large enrichment of genes of its own.

In situ hybridizations of example genes (e.g., some dopamine and glutamate receptors) revealed that most of the song nuclei expression patterns were consistent with the brain subdivisions to which they belonged (Fig. 3, A to C, and table S2) (*32*–*34*). However, as seen previously (*33*, *35*, *36*), some of the song nuclei had highly differential expression from their surrounding brain divisions (i.e., *FMNL1*, *DGKI*, and *GPSM1* in Fig. 3, A to C). The most song-nucleus–specific gene was *FAM40B* (also called *STRIP2*), a phosphatase that was restricted to cortical-like song nuclei and the primary cortical sensory populations (like auditory area L2 in Fig. 3A).

A dendrogram analysis separated the cortical song nuclei from the striatal and showed a stronger relationship between HVC and LMAN of the nidopallium (Figs. 2B and 1A), consistent with the recently revised understanding of avian brain organization and homologies with mammals (*5*, *6*, *37*). These findings show that even before singing starts, the song-learning nuclei have thousands of differentially expressed genes that define specific molecular functions for each [see (*15*) for characterization of the specializations in song nuclei].

Singing activates both a core and regionally diverse patterns of genes

Of the 24,498 transcripts, we found an estimated 2740 (~11%) that were singing-regulated, up or down in time, in one or more song nuclei (Fig. 4, A and B, and table S8). These transcripts mapped to 1833 genes, indicating a conservative estimate of ~10% of the transcribed avian genome that is regulated by singing behavior. Area X had the most regulated transcripts (1162), followed by HVC (772), RA (702), and LMAN (635) (Fig. 4B) (the sum is higher than 2740 because of transcripts expressed in more than one song nucleus). A small number of genes (82) had singing-regulated splice variant differences (table S9), consistent with splice variant differences at baseline among song nuclei for glutamate receptor subunits (*33*), which can regulate activity-dependent genes in the brain. The vast majority (96%) of the 2740 singing-regulated transcripts were enriched in only one or two song nuclei, and a core set of only about 97 transcripts was regulated in at least three or four (<1.0%) song nuclei; of the latter, only 20 genes were equally regulated in all four song nuclei (Fig. 4, A and B, and table S8, green and yellow).

The core set of 97 transcripts was enriched for known IEGs (*38*), including membrane depolarization–regulated (Ca^{2+} responsive) genes identified in cultured hippocampal (*39*) and cortical neurons (*40*) and genes induced in the auditory pathway by hearing song (*41*) (tables S10A and S7). In contrast, the brain region–specific singing-regulated genes had very little overlap with classic IEGs or a list of cell cultured–defined depolarization-induced genes (table S10A). Rather, the striatal Area X singing-regulated genes were enriched for cytoskeletal neural connectivity and neural migration functions, and RA was enriched for mitogen-activated protein kinase pathway transcripts, which control gene expression, differentiation, and cell survival. This suggests that our in vivo analyses are useful for finding region-specific or stimulus-specific genes that may be relevant for the underlying singing behavior.

Similar to the baseline expression, in situ hybridizations revealed that song nuclei expression patterns were consistent with the brain subdivisions to which they belong (Fig. 3, A to C, and table S3), except that the surrounding brain areas in some birds tended to have lower expression, presumably because they sang without much other movement behavior to cause movement-induced gene expression in the surrounding regions (*42*). We also noted that even among the core early-response genes induced in all song nuclei, expression levels at baseline differed among song nuclei (Fig. 3D). This suggests that there is even greater diversity among the song nuclei singing-regulated genes than simply presence or absence of regulation.

Analysis of the behaviorally regulated gene expression across time, using unsupervised hierarchical clustering (SM8), revealed up to 20 temporal profiles (clusters) among the four song nuclei, including transient or sustained, increased or decreased, early (0.5 to 2 hours) or late (3 to 7 hours), or two peaks of expression (fig. S3, A to D, and table S8). These 20 clusters can be further grouped into four superclusters

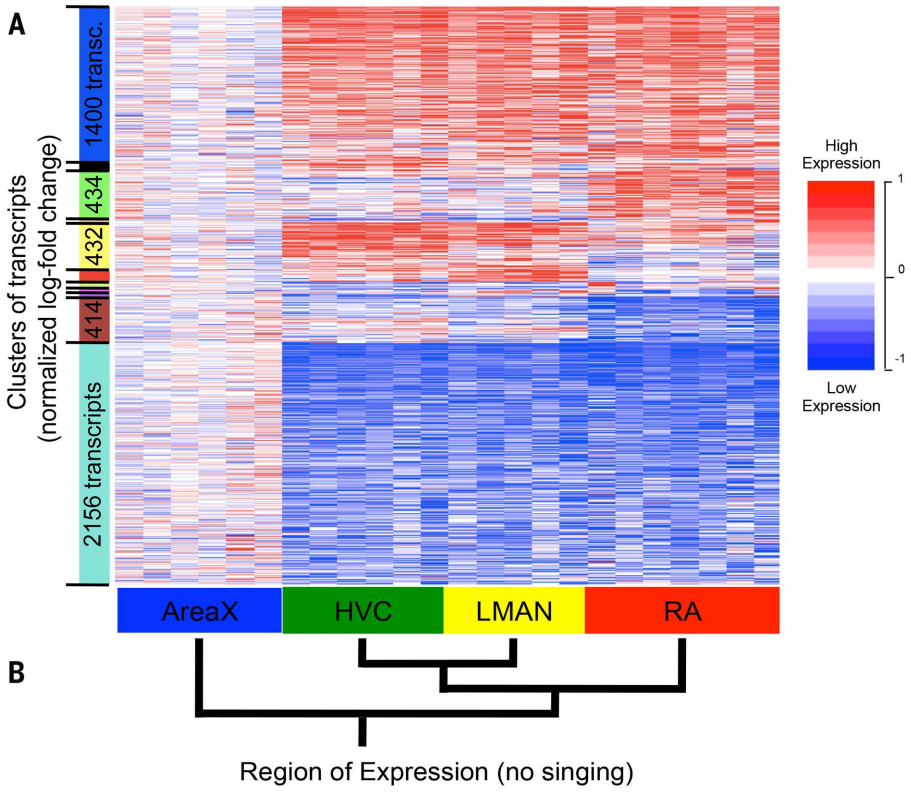


Fig. 2. Region-enriched gene expression at baseline. (**A**) A heat map of hierarchically clustered expression profiles of 5167 transcripts (rows) that are differentially expressed across regions at baseline (FDR $q < 0.1$; see fig. S11 for FDR $q < 0.2$) in silent birds (red, increases; blue, decreases; white, no change) relative to mean Area X expression (numbers of transcripts not shown for small clusters). Each transcript is normalized to the average value of expression in Area X. Each column is an animal replicate. Detailed results are in table S4. (**B**) Average linkage hierarchical tree, generated from mean expression in each brain region, representing the molecular expression relationships between regions.

of temporal profiles: (i) transient early increases, (ii) late-response increases, (iii) transient early decreases, and (iv) late-response decreases (Fig. 5, A to D). Only three of the temporal clusters had relatively comparable representations of genes in all brain regions, all belonging to transient early-increase clusters, including the IEG 0.5 to 1 hour cluster (Fig. 5A; fig. S3, tan cluster; and table S11), which contained a significant proportion (16%) of the core set of 97 transcripts ($P < 1 \times 10^{-5}$, hypergeometric test). For the remaining supertemporal profiles, each song nucleus had a region-enriched set of genes, except the late-response increasing pattern in LMAN (Fig. 5, fig. S3E, and table S11).

Functional enrichment analyses showed that the activity-regulated gene expression sets from previous cell culture experiments (table S7) were highly enriched in the early transient IEG temporal cluster expressed in all song nuclei (table S10B). All of the late-increase singing-regulated clusters (Fig. 5B) also had detectable functional enrichments of genes, with Area X+HVC enriched in calcium ion binding and phosphatase proteins (blue temporal cluster); Area X late-increase genes were additionally enriched in chromosome organization, biogenesis (green), activity-dependent late-response genes identified in cultured neurons (*40*) (turquoise), and ribosomal proteins (black); HVC was additionally enriched in RNA-protein complexes and PSD proteins (cyan); and RA late-increase genes (salmon) were enriched in a different set of calcium ion–binding and ribosomal proteins (table S10B and Fig. 5B). Notably, we did not find any functional enrichment for the remaining transiently increased clusters or any of the decreased clusters, except genes regulated by the serum response transcription factor (SRF) in the slow decreasing cluster of RA (table S10B and Fig. 5D, yellow). These findings show that all song nuclei share a core set of genes with rapid transient up-regulation, but each song nucleus has its own dominant (though partly overlapping) set of other early- and late-responsive behaviorally regulated genes, suggesting cascades of gene regulation specific to each song nucleus with functions that remain to be discovered.

Relationships between differential baseline and differential singing-regulated genes

We next investigated how a small core set of behaviorally regulated transcription factors expressed in most brain regions could regulate a diverse set of downstream genes, with little overlap among regions. We hypothesized that the differential transcriptional state at baseline, before cell stimulation with singing, affects region-enriched singing-regulated expression (*43*, *44*). Three lines of evidence support this hypothesis. First, hypergeometric tests revealed significant overlap between subsets of transcripts from the baseline region-enriched clusters (Fig. 4C, top gray box) with the singing-regulated region-enriched clusters (Fig. 4C, red lines and table S12) and with 10 of the 20 temporal clusters (Fig. 4C, blue and black lines between two gray boxes). If a gene was expressed at higher levels in a region relative to others at baseline before singing, it was also more likely to increase in that region during singing; the converse was not true for the decreasing sets of singing-regulated genes.

Second, a genome-wide binding site analysis of motifs for transcription factors (SM11) (*45*, *46*) revealed ~100 motifs enriched in regulatory regions (e.g., directly upstream of transcription start sites) of genes in the temporal behaviorally regulated clusters (tables S13 and S14 and Fig. 6, A and B), and these matched genomic locations were also found in mammalian genomes (*47*, *48*). With these motifs, we performed an association analysis between the region-specific and temporal clusters of genes to generate song nuclei–specific transcription factor motif to gene cluster networks (Fig. 6C, simplified network; fig. S4, detailed network; and table S15, edge list) [statistical significance tested with Euclidean distance to randomly generated networks (SM11 and SM12)]. Consistent with the core IEG cluster findings, we found that binding sites for five early-activated transcription factors (EATFs) (*MEF2*, *SRF*, *NFKB*, *CREB*, and *CaRF*) that are constitutively expressed at baseline and activated in response to neural activity (*38*, *49*, *50*) were significantly overrepresented in the singing-regulated cluster of IEGs expressed in most song nuclei (Fig. 6C and figs. S4 and S5A). In turn, the binding motifs of the singing-regulated *AP-1* (bound by a FOS-JUN dimer) and *EGR1* IEG transcription factors were also enriched directly upstream of the transcription start sites of many genes in our avian IEG cluster (Fig. 6, A to C). *EGR1* can bind to its own promoter and down-regulate itself (*51*), which is consistent with the transient increase and subsequent decrease of some transcripts in the IEG temporal cluster. Also overrepresented in the IEG cluster was the *ARNT* motif, which also has the binding motif for the IEG *NPAS4*.

Third, consistent with our region-specific clusters, some transcription factors that were differentially expressed in a region or a combination of regions at baseline had binding motifs in genes that were differentially regulated in that region(s) at baseline or during singing. For example, variants of the *NFE2L1* and *MAF* transcription factors that dimerize and bind to the *TCF11* motif (*52*) were higher or lower in Area X relative to the pallial song nuclei at baseline (fig. S6), and

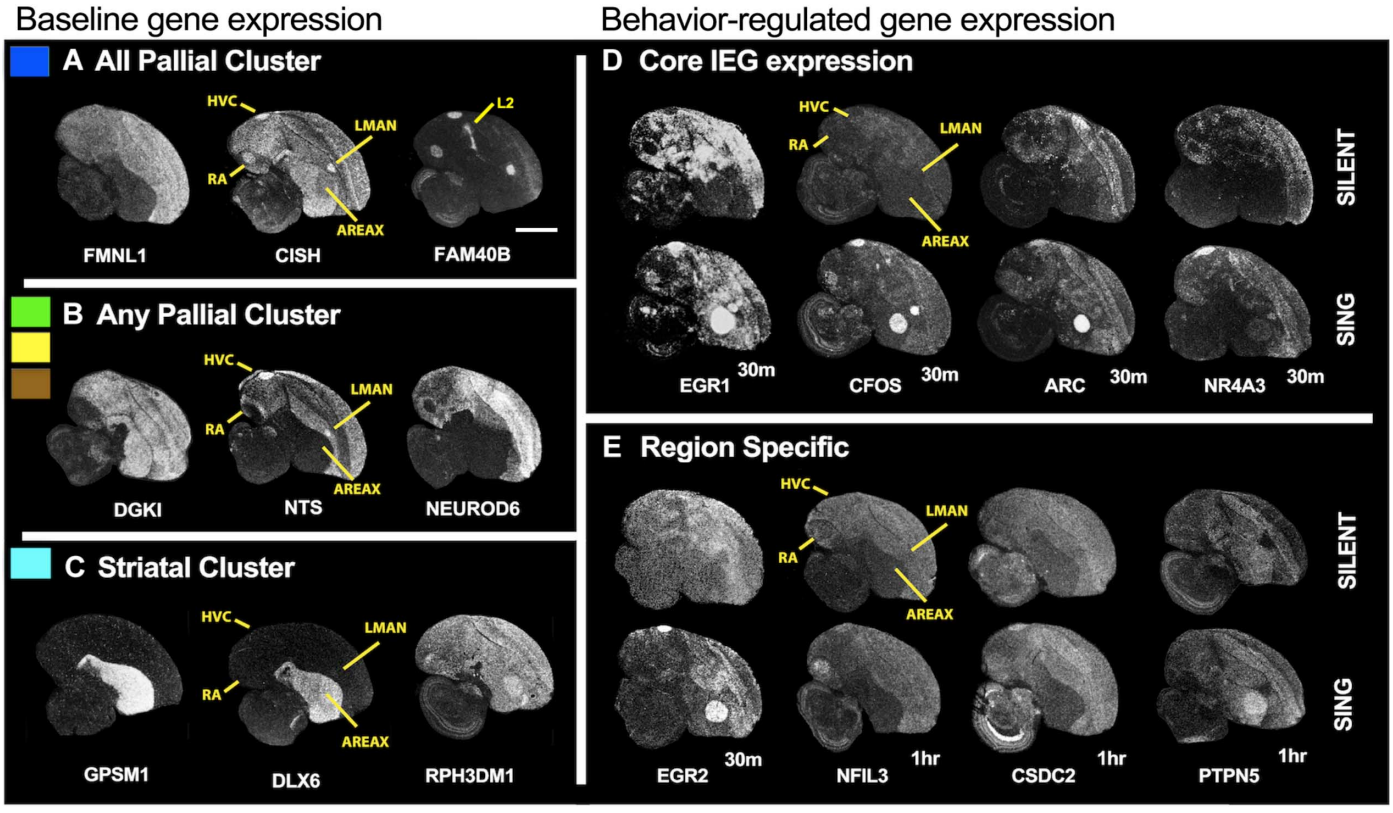


Fig. 3. In situ hybridizations of baseline and singing-regulated genes. (**A**) Genes higher in all pallial song nuclei (RA, HVC, and LMAN) relative to the striatal song nucleus (Area X) at baseline (Fig. 2A, blue clusters). (**B**) Genes differentially expressed just among the pallial song nuclei (green, yellow, and brown clusters) at baseline. (**C**) Genes higher in the striatal song nucleus relative to pallial song nuclei (turquoise cluster). (**D**) Core singing-regulated genes regulated in three to four song nuclei detected by microarrays but detected in all four with diverse levels by in situ hybridization, most peaking at 30 min. (**E**) Region-enriched singing-regulated genes in one or two song nuclei, with peaks of expression at later time points. Film autoradiograph images are inverted, showing white as labeled mRNA expression of the gene indicated below the image. Dorsal is up; anterior is right. Scale bar, 2 mm.

the *TCF11* binding motif was overrepresented in the slow-increase singing-regulated cluster of genes in Area X (Fig. 6C and figs. S4 and S5B). However, there were many other cases where EATFs and other transcription factors did not exhibit differential regional baseline expression but had binding motifs enriched in clusters of singing-regulated genes specific for a song nucleus. For example, the EATF transcription factors *SRF* and *CaRF*, which are not differentially expressed at baseline (table S5), had strong motif associations to singing-regulated genes in Area X and HVC. The *MZF1* and *PRRX2* transcription factors had associations with different sets of genes in Area X and RA (Fig. 6C and figs. S4 and S5B). Thus, we experimentally tested whether one of these EATFs, *CaRF*, regulated the predicted region-specific genes (Fig. 7).

CaRF is required for regulation of both core and regional expressed sets of genes

We investigated the Ca^{2+} responsive transcription factor *CaRF* because the network analyses implicated it in both the regulation of the Ca^{2+} responsive IEGs that are induced in most song nuclei and some that are regionally enriched in Area X and HVC (Fig. 4C and fig. S6). Because we lacked an established zebra finch neural cell culture method to test *CaRF* function, we used RNA interference (RNAi) against *CaRF* in cultured mouse cortical neurons and hybridized labeled cDNA to mouse oligonucleotide microarrays representing many of the same genes on our zebra finch oligonucleotide microarray (SM4). We identified a set of genes that showed decreased or increased expression after *CaRF* knockdown independent of membrane depolarization (Fig. 7A and table S16), and many of these function in calcium signaling pathways (fig. S7 and table S17) (*53*). This is consistent with the proposed role of *CaRF* in regulating neuronal gene expression under basal neural activity (*48*, *54*), as both a repressor and activator (*48*). Importantly, as predicted by our promoter motif analyses in birds, the ranked list of *CaRF*-regulated genes showed enrichment for singing-regulated genes that had a nearby *CaRF* binding site (P = 0.0014, Wilcox test) (Fig. 7B). This enrichment was highest in the set of genes regulated in Area X and HVC (Fig. 7B), supporting our network result (Fig. 6C).

CaRF RNAi knockdown also caused genes that were normally up-regulated by membrane depolarization to be suppressed to normal baseline levels and, conversely, genes that were normally down-regulated by membrane depolarization to be up-regulated (Fig. 7C and table S18). This suggests that *CaRF* is required to buffer activity of these gene promoters under basal conditions such that they can become stimulus-responsive upon membrane depolarization. Importantly, this same set of membrane depolarization- and *CaRF*-regulated genes significantly overlapped with those that had the *CaRF* binding site in the singing-regulated genes of the IEG (tan) cluster. They also significantly overlapped with several other clusters that were specifically up-regulated in Area X and HVC (Fig. 7D, magenta and cyan clusters; table S19; and fig. S3E). Genes that showed decreased expression preferentially in RA, but also in other song nuclei (fig. S3, yellow), after 2 to 3 hours of singing (the same amount of time the cultured cells were depolarized) had even greater overlap (Fig. 7D, yellow).

Overall, the findings demonstrate a requirement of the *CaRF* transcription factor for baseline and activity-dependent regulation of some of the very same genes for which we found *CaRF* binding motifs that are regulated at baseline and by singing in a region-specific manner, respectively. The calcium signaling and calcium ion–binding genes tended to increase during song production and were affected in the *CaRF* knockdown experiments, which is evidence of consistent *CaRF* function across species. We next sought an explanation of how EATFs that are

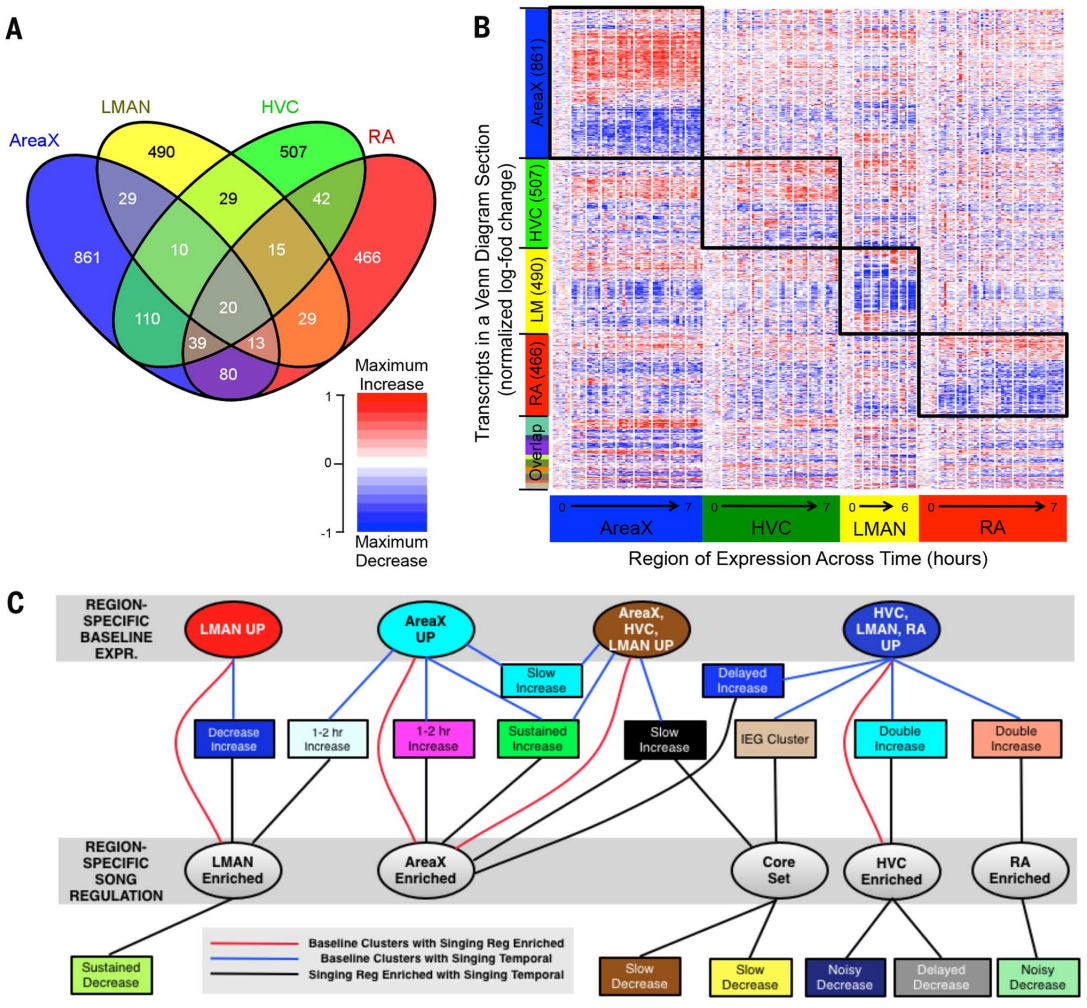


Fig. 4. Region-enriched gene expression in response to singing. (**A**) A four-way Venn diagram showing regional singing-regulated distribution of 2740 transcripts (FDR q < 0.2). (**B**) Heat map of all 2740 transcripts from the Venn diagram, hierarchically clustered independently in all four song nuclei, then sorted by increased or decreased expression, and level of significance from highest to lowest in the linear model. Each column (170 total) is an animal replicate within a time point, and white lines separate time points. Red, increases; blue, decreases; white, no change relative to 0-hour samples for each song nucleus. Each transcript is normalized so that the maximum increase relative to nonsinging birds in any region is the darkest shade of red for increasing transcripts, and the maximum decrease is the darkest shade of blue for decreasing transcripts. Boxes highlight significant behaviorally regulated enrichment for each region (FDR q < 0.2 for that region). Figure S12 shows a more stringent heat map of region-enriched expression with a similar result. (**C**) Relationships among clusters of transcripts from the baseline region-enriched (top gray box, from Fig. 2A), singing temporal-enriched (rectangular nodes, from fig. S3, A to D), and singing region-enriched [bottom gray box, from (B)] patterns. Nodes are colored according to their cluster colors in the respective figures. Edges between two nodes correspond to significant overlap between two groups of transcripts (P < 0.001, hypergeometric test). Nodes are sorted to optimize noncrossing of edges. Detailed results are in table S8.

not differentially expressed at baseline could regulate these genes in a region-specific manner.

Epigenetic modifications predefine region specificity of gene regulation

Although transcription factors are the ultimate regulators of gene expression, their ability to bind to sites in the genome is gated by chromatin structural changes. Chromatin regulation by acetylation of histone 3 at lysine 27 (H3K27ac) has been extensively studied and shown to be a strong indicator of active enhancers (*55*). We thus performed an experiment to identify active transcriptional regulatory regions in the genomes of individual dissected song nuclei (RA and Area X, which showed the largest regional differences) before and after singing, as measured by a genome-wide histone ChIP-seq analysis of H3K27ac (SM14, SM15, and table S20). The active genomic regions can be searched as tracks in the University of California–Santa Cruz (UCSC) browser against the zebra finch genome (*56*). This analysis also required that we create a more stringent selection of regional, early, and late singing-responsive genes from the respective clusters in RA and Area X (Fig. 5 and fig. S3), using principal components analyses (fig. S8).

Out of 35,958 peaks, we found 30% (10,749) enriched in Area X and 21% (7673) enriched in RA. Under basal conditions, genes with song nuclei–specific expression patterns had nearby genomic regions that were significantly more likely to be marked by H3K27ac in that brain region (Fig. 8A, blue and red, and table S21) (~1300 genes). Conversely, genes that were expressed similarly in RA and Area X did not show a significant regional bias in the distribution of this chromatin mark (Fig. 8A, gray, and table S21) (~1100 genes examined). Interestingly, when we considered only the set of RA or Area X region-specific genes that were also up-regulated by singing, we found that they were already associated with higher nearby H3K27ac in their preferred brain region before singing (Fig. 6, B, D, and E; fig. S9, A to E; and table S22). There was a strong positive correlation between differences in nearby H3K27ac at baseline and differences in singing-dependent up-regulation of these genes in RA and Area X (R = 0.37, P = 1.6×10^{-12}; Pearson correlation). Conversely, late-response genes that were comparably induced by singing in both RA and Area X showed comparable H3K27ac under basal conditions (Fig. 8B, gray, and table S22). Furthermore, the early-response cluster of genes, which were expressed and induced comparably in both RA and Area X (e.g., *FOS*), also showed comparable H3K27ac in both brain regions at baseline (Fig. 8C and figs. S9A and S10A). Notably, we did not find any significant difference [e.g., 0 significant peaks; false discovery rate (FDR) threshold < 0.01] in H3K27ac peaks within either song nucleus when we compared ChIP-seq profiles obtained before and after singing (fig. S10A). We detected a weak signal for increased H3K27ac peaks in the Area X down-regulated genes (fig. S10B).

These data suggest that the regional differences in chromatin activity present before singing

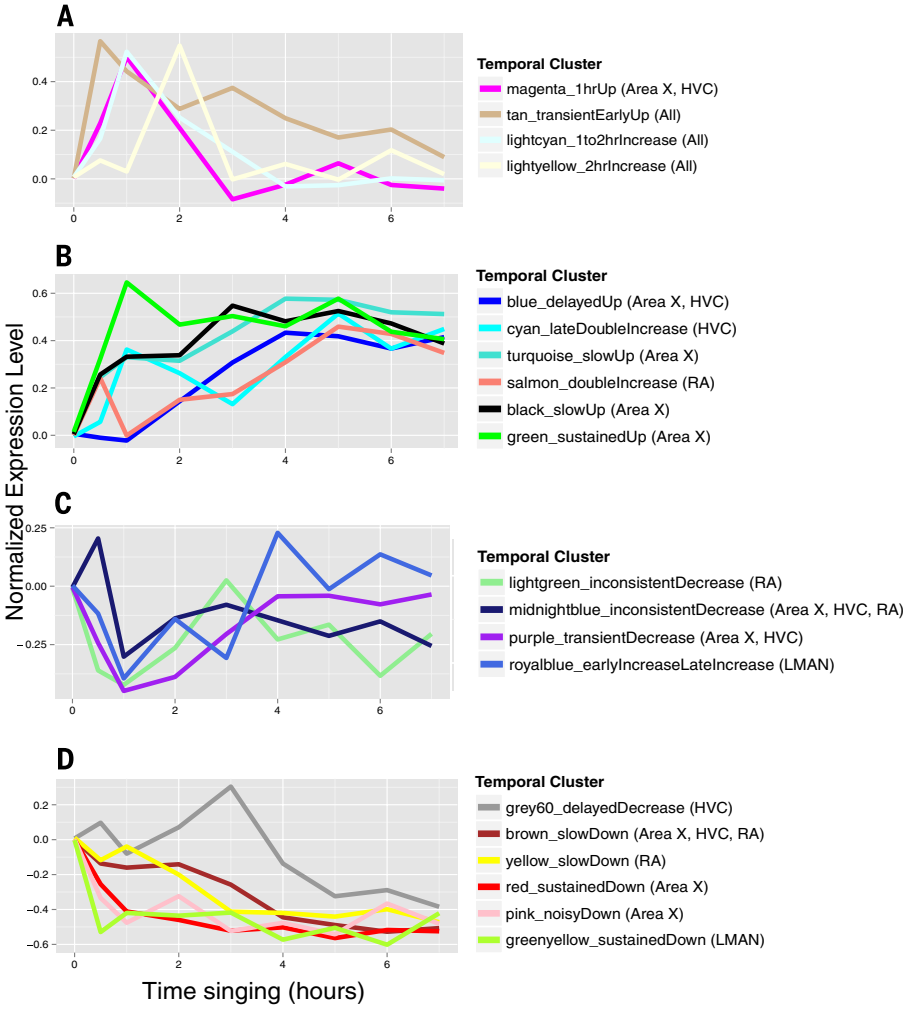


Fig. 5. Temporal singing-regulated patterns across time. (**A**) Averages of gene expression levels in four temporal clusters of transient early response increases. (**B**) Averages of six late-response gene cluster increases. (**C**) Averages of four transient early-response cluster decreases. (**D**) Averages of six late-response gene cluster decreases. The temporal profiles are normalized such that nonsinging birds have a value of 0 and each gene has a maximum increase or decrease of 1. Each point represents the mean across all gene-brain region combinations for that time point. The 20 colors match the major temporal clusters in fig. S3, A to D.

begins are predictive of differential singing-dependent induction of late-response genes. This hypothesis was further supported by our observation of regional H3K27ac differences at baseline for 50 genes that had equivalent basal expression in RA and Area X but region-specific up-regulation upon singing (table S22, blue and red highlights). An ingenuity pathway analysis on the Area X set of genes out of the 50 mentioned above (table S22, blue, and SM15) revealed that they were enriched for locomotion behavior (P = 0.004; *ARNTL*, *CALB1*, *FGF14*, *RCAN2*, and *RIMS1*) and movement-disorder functions (P = 0.004; *ARNTL*, *CALB1*, *CAPZB*, *DIRAS2*, *EEF1A2*, *ELMO1*, *FGF14*, *MTMR2*, *RPSA*, and *TMED10*), consistent with the function of Area X and the surrounding striatum. There were too few RA-specific genes without baseline differential expression (10 genes) to be tested by pathway analyses. Overall, these findings indicate that region-specific epigenetic chromatin activity at or near transcription factor binding sites for transcription factors expressed in all brain regions could determine which singing or baseline differentially regulated genes are expressed in each brain region.

Discussion

The magnitude of the anatomical diversity of behaviorally regulated genes and their networks in different brain regions of the same circuit was unexpected (*24*, *29*, *30*, *41*). Our findings suggest two mechanisms that control this diversity: (i) region-enriched transcription factors that regulate region-enriched expression of their target genes and (ii) region-enriched epigenetic marks that determine which genes can be expressed in specific brain regions in both baseline and behaviorally regulated states. The first mechanism is consistent with the hypothesis that interactions between early transcription factors and late-response genes coordinate activity-dependent gene induction associated with behavior (*57*) but, in this case, in a region-specific manner. The second, epigenetic, mechanism is just beginning to be explored at the level of neural activity (*40*, *58*) and has not been addressed in complex behaviors.

Given our findings and known signaling pathways from experiments in cultured cells (*59*), we propose the following overall mechanism (see the figure in the print summary, page 1334). Neural activity during the performance of a behavior, such as singing, causes release of neurotransmitters at the synapses between connected cells and activates postsynaptic receptors. These receptors initiate an intracellular signaling response that alters the activity, often through phosphorylation, of constitutively expressed EATFs. The activated EATFs bind or are already bound to the open chromatin of promoters or enhancers of the core IEGs enabled in all brain regions, as measured by H3K27ac, to activate their expression. The IEGs in turn, along with EATFs, bind to recognition regions of open chromatin that have already been primed in a cell type–specific manner, which leads to the induction of region-specific late-response genes. Some transcription factors are already expressed in a region-specific manner and add to the diversity of regulation of the downstream genes. Furthermore, our data show that brain region–specific open enhancers or promoters are already waiting in an active state, ready to do their job at a moment's notice when the neurons fire to turn on programs of gene expression. Thus, the production of learned behavior modulates an already primed transcriptional and epigenetic network specific to different subregions of the circuit that controls the behavior.

This model may be an explanation for the finding that the IEG and EATF *NPAS4*, in response to neural activity, activates different sets of genes in cultured excitatory versus inhibitory neurons (*60*). Likewise, we find that common induction of IEGs across the many different kinds of neurons that comprise all song nuclei is associated with distinct programs of late-response genes, which are likely dependent at least in part on IEG regulation. However, one notable difference between our data and a recent study of activity-dependent enhancers in cultured neuron preparations is that, whereas membrane depolarization was found to further induce H3K27ac at enhancers near activity-regulated genes (*58*), we find that H3K27ac peaks in vivo in the brain are already enriched near singing-inducible genes under basal conditions and do not show further activation upon singing. It is possible that the neural networks recruited upon singing are sparse enough in the song nuclei that we were unable to detect H3K27ac changes in these cells against the background noise. An alternative possibility is that ongoing neural activity in the brain of an awake behaving animal is sufficient to keep enhancers poised in a fully active state even before execution of a specific behavioral task like

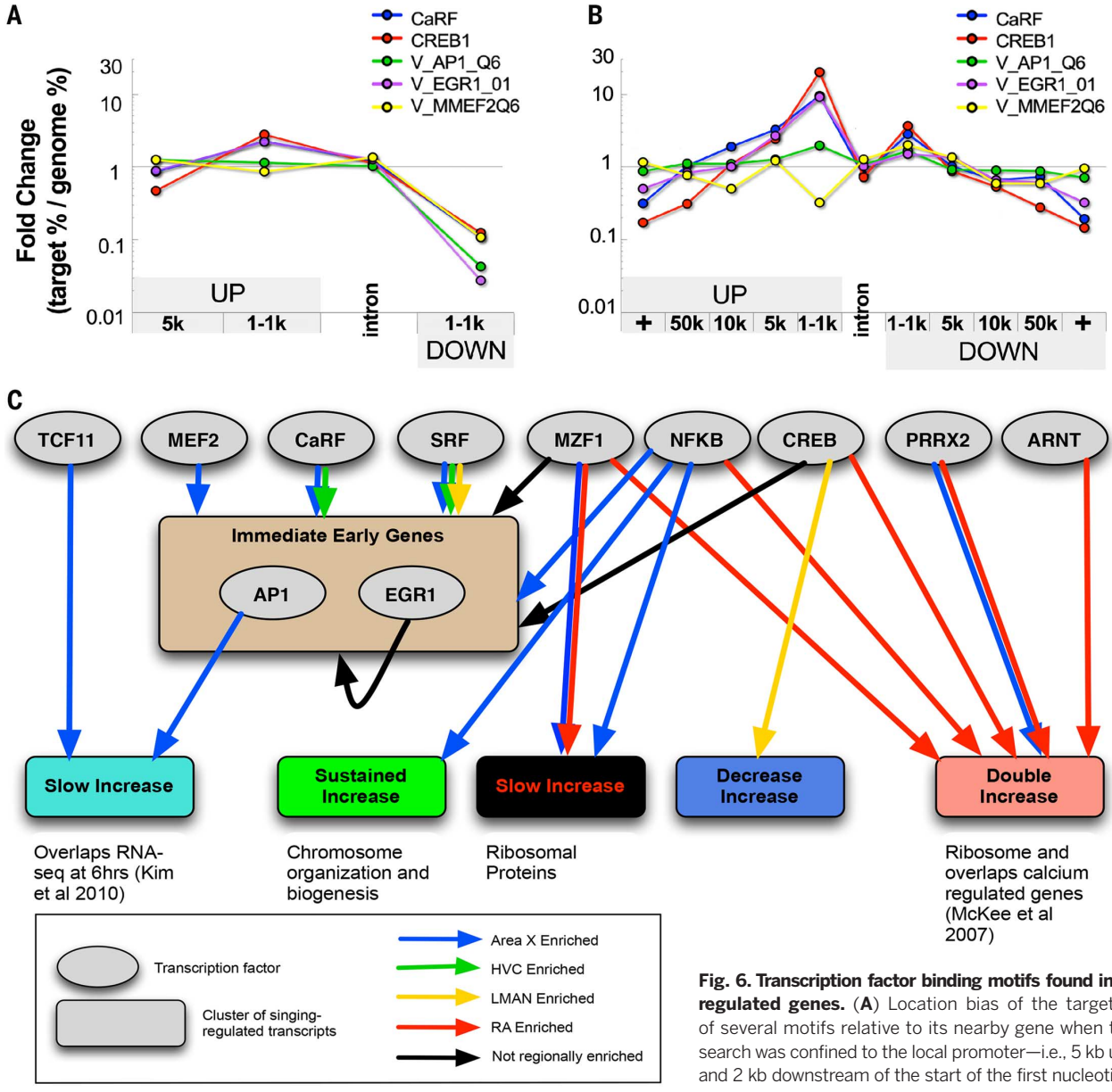


Fig. 6. Transcription factor binding motifs found in singing-regulated genes. (**A**) Location bias of the target window of several motifs relative to its nearby gene when the motif search was confined to the local promoter—i.e., 5 kb upstream and 2 kb downstream of the start of the first nucleotide of the first exon of the gene. Fold change (plotted on the log scale *y* axis) is the ratio of the percentage of the motif target windows that fell within a particular position category relative to the first exon of a gene (target %) versus the percentage of windows that fall within that position category genome-wide (genome %). (**B**) Location bias of the motif target window relative to its nearby gene when the motif search was performed over the gene territory—i.e., halfway upstream and halfway downstream to the last or first exon of the nearest nonoverlapping gene. (**C**) Transcription factor motif-gene cluster network summarized from fig. S4 showing relationships between enriched EATFs (gray circles) and their binding motifs in subsets of genes from the temporal singing-regulated clusters (colored rectangular nodes as in fig. S3, A to D). Edges are colored on the basis of the region-specific expression of the predicted regulatory targets of the TF within each singing-regulated cluster (SM11 and SM12). Detailed results are in table S13 and fig. S4.

singing. In this model, it is regulation of sequence-specific DNA binding of transcription factors that is most important for instructing the level and nature of gene expression, whereas epigenetic marks on chromatin are permissive for expression of the predetermined program.

Our *CaRF* manipulation experiments help reveal further complexity and potential novel mechanisms of activity-dependent gene regulation in the brain. The increased activity-regulated genes that are reversed in the absence of *CaRF* in response to membrane depolarization suggest that *CaRF* may act as a modulating transcription factor for neural activity–dependent regulation of its target genes. In this scenario, it prevents differential expression of its target genes until neural firing increases. When *CaRF* is removed by knockdown, it can no longer buffer the expression of these genes in the absence of activity; consequently, in the presence of activity, other factors can regulate the genes in a direction opposite of what *CaRF* would do. The specific mechanisms by which *CaRF* might achieve this function remain to be determined, but the H3K27ac enhancer activity in *CaRF* target genes is likely to play a role.

Additional transcriptional anatomical diversity not tested in this study could possibly be generated with differential expression of neurotransmitter membrane receptors at baseline in different brain regions, which could activate different signaling pathways in those neurons during singing (*2*, *33*). Our hypothesis does not explain the down-regulation of some gene clusters where regionally specific transcription factor motifs were not enriched in those genes, and

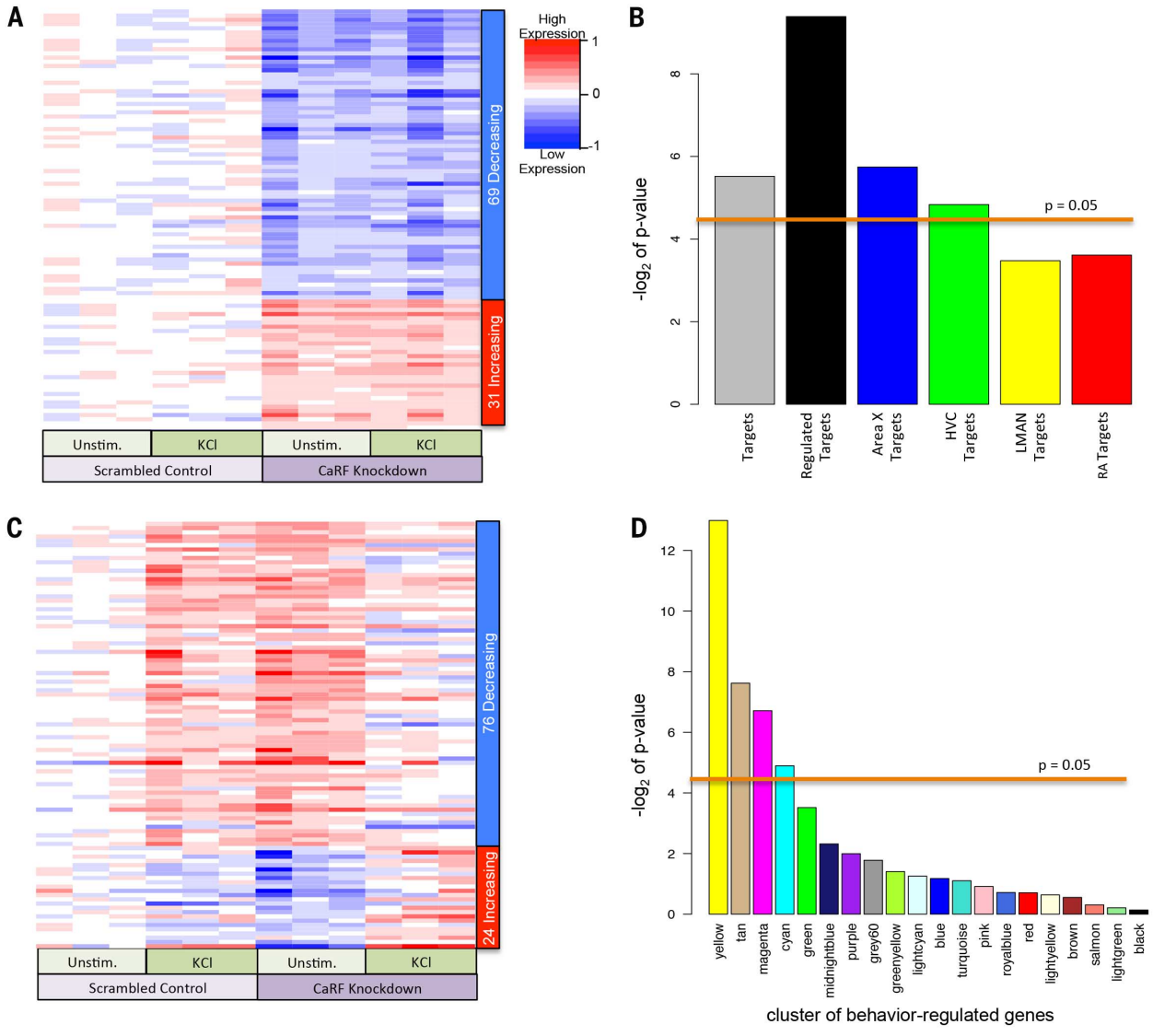


Fig. 7. RNAi knockdown illuminates *CaRF* binding motif relationships with singing-regulated genes. (**A**) Heat map of genes affected by *CaRF* knockdown independent of membrane depolarization in mouse cultured neurons. Rows represent the 100 transcripts most changed by *CaRF* RNAi knockdown ($P < 0.0014$; FDR $q < 0.475$), sorted according to the *t* statistic, which takes direction of regulation into account. Each column is an independent sample ($n = 3$ unstimulated controls; $n = 3$ KCl depolarized in the presence of either scrambled RNAi or *CaRF* RNAi knockdown virus). Color intensities (blue to red) represent the log fold change in knockdown cells relative to the mean of the scrambled control conditions. (**B**) Significance of the enrichment of zebra finch baseline genes (cluster colors according to Fig. 2A) with *CaRF* promoter motifs in the ranked list of *t* values for *CaRF* knockdown–affected genes in mouse cultured neurons. $P < 0.05$ (above line) is a significant association, Wilcox rank sum statistic over multiple permutations (*66*). (**C**) Similar to (A), except for genes that respond differently to KCl activity in the *CaRF* knockdown cells. Rows represent the 100 transcripts most changed in expression ($P < 0.015$, factorial test), sorted according to the *t* statistic. (**D**) Significance of the enrichment of zebra finch singing-regulated genes (cluster colors according to Fig. 5 and fig. S3), with *CaRF* promoter motifs in the ranked list of *t* values for genes differentially regulated by neural activity in mouse cortical neurons during *CaRF* knockdown versus control. $P < 0.05$ (above line) is a significant association, Wilcox rank sum statistic over multiple permutations (*66*).

thus their regulation would have to be explained by other mechanisms.

Our findings suggest that each song nucleus has diverse molecular functions and gene networks. Consistent with their dominant roles in song production (*7*–*13*) compared to other song nuclei, HVC is specifically enriched with singing-regulated increases in PSD proteins used for cell-to-cell communication and RNA-protein complexes, and RA is enriched with genes in the mitogen-activated protein kinase (MAPK) pathway, such as *DUSP1*, which is proposed to be involved in neural protection of a brain region that is highly active during behavior performance (*61*, *62*). Consistent with their dominant roles in learning (*7*–*13*), LMAN shows greater specificity for the cAMP response element–binding protein (CREB) pathway, a key transcription factor involved in learning and memory (*59*, *63*), and Area X is more enriched with expression of neural connectivity, chromosome organization, and biogenesis genes. In addition, the large overrepresentation of noncoding RNA genes expressed at baseline in Area X indicates that its transcriptional regulatory network may be more extensive than the pallial song nuclei. The larger overrepresentation of neural connectivity and cell signaling genes in the pallial song nuclei indicates greater focus on cell structure and communication.

In terms of memory, a long-held hypothesis is that neural activity will induce an early wave of responsive genes, which in turn regulate a late wave of genes, and that the first wave would act as a molecular switch converting short-term memories into long-term memories (*57*, *64*, *65*).

If true, singing would be associated with continuous memory consolidation and song fine-tuning, with each nucleus having specific waves of gene regulation for their specific functions. An alternative, not mutually exclusive, proposal states that the activity-dependent waves function as a metabolic mechanism to maintain protein turnover for normal cell homeostasis due to increased protein catabolism that occurs during high activity levels (*17*). If true, it would be associated with continued repair of the circuit when used. Our transcription factor binding motif

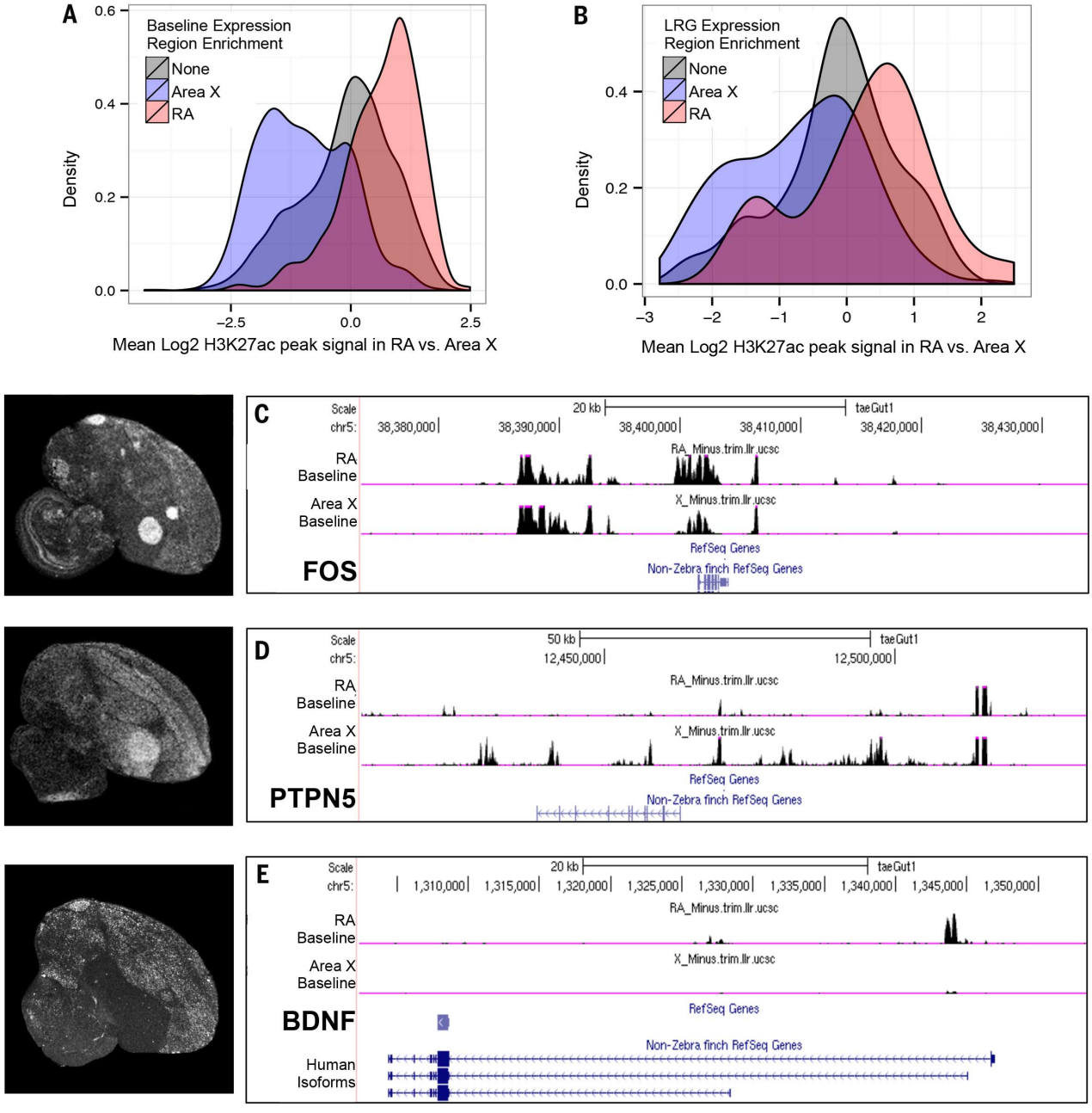


Fig. 8. Region-specific epigenetic signatures predefine behaviorally regulated gene expression. (**A**) Density plot of genes differentially expressed at baseline in RA versus Area X and the difference in the level of nearby H3K27ac peaks in the genomes of cells in RA versus Area X. Each H3K27ac peak is mapped to a gene with the nearest transcription start site. For each gene, the changes in all mapped H3K27ac peaks are averaged. The H3K27ac distributions for RA versus Area X enriched genes are significantly different ($P = 1.5 \times 10^{-186}$, *t* test). (**B**) Similar plot as in (A) except for differentially expressed late-response singing-regulated genes. The distributions for RA and Area X are also significantly different ($P = 1.8 \times 10^{-5}$, *t* test). However, there are two peaks in RA, which suggests that active genomic sites in Area X in the negative peak for RA could be genes that are actively suppressed in Area X. Corresponding data can be found in tables S21 and S22. (**C**) H3K27ac peaks surrounding a gene induced by singing across all brain regions, *FOS*; (**D**) H3K27ac peaks of a gene induced specifically in Area X, *PTPN5*. (**E**) H3K27ac peaks of a gene induced at low levels in RA but not detectable in Area X, *BDNF*. The plots show the log-likelihood ratios of H3K27ac signal in pooled baseline RA and pooled baseline Area X samples versus input DNA around the genomic regions in the zebra finch. The relevant gene models from the UCSC genome browser are shown below. Peaks measure both enhancer and promoter regions. Left of the H3K27ac peaks are in situ hybridization mRNA signal in singing animals. FOS and PTPN5 are shown in Fig. 3, and BDNF is used with permission from (*37*).

analysis suggests that both the early and late transcriptional responses could be driven by some of the same EATFs. This would indicate that the two waves of gene expression may not entirely depend on each other and that they could be used for both memory and homeostasis functions.

In summary, as the mechanisms that define the genome-phenotype relationship, including the diversity of gene expression patterns, begin to be understood, so will the role of individual genes and pathways in learning, maintenance, and production of behavior. Performance of complex behavior involves interaction between neural activity, networks of cells, and networks of genes. Untangling the subtle differences in connected neurons, firing patterns, signaling pathways, and transcription factor activity may lead to a greater understanding of the diversity of the gene expression patterns that we observe here in highly interconnected cells within an intact multicellular organ.

REFERENCES AND NOTES

1. A. J. Doupe, P. K. Kuhl, Birdsong and human speech: Common themes and mechanisms. *Annu. Rev. Neurosci.* **22**, 567–631 (1999). doi: 10.1146/annurev.neuro.22.1.567; pmid: 10202549
2. E. D. Jarvis, Learned birdsong and the neurobiology of human language. *Ann. N. Y. Acad. Sci.* **1016**, 749–777 (2004). doi: 10.1196/annals.1298.038; pmid: 15313804
3. R. Mooney, M. Konishi, Two distinct inputs to an avian song nucleus activate different glutamate receptor subtypes on individual neurons. *Proc. Natl. Acad. Sci. U.S.A.* **88**, 4075–4079 (1991). doi: 10.1073/pnas.88.10.4075; pmid: 11607180
4. M. Luo, D. J. Perkel, A GABAergic, strongly inhibitory projection to a thalamic nucleus in the zebra finch song system. *J. Neurosci.* **19**, 6700–6711 (1999). pmid: 10414999
5. A. Reiner *et al*., Revised nomenclature for avian telencephalon and some related brainstem nuclei. *J. Comp. Neurol.* **473**, 377–414 (2004). doi: 10.1002/cne.20118; pmid: 15116397
6. E. D. Jarvis *et al*., Avian brains and a new understanding of vertebrate brain evolution. *Nat. Rev. Neurosci.* **6**, 151–159 (2005). doi: 10.1038/nrn1606; pmid: 15685220
7. F. Nottebohm, T. M. Stokes, C. M. Leonard, Central control of song in the canary, Serinus canarius. *J. Comp. Neurol.* **165**, 457–486 (1976). doi: 10.1002/cne.901650405; pmid: 1262540
8. S. W. Bottjer, E. A. Miesner, A. P. Arnold, Forebrain lesions disrupt development but not maintenance of song in passerine birds. *Science* **224**, 901–903 (1984). doi: 10.1126/science.6719123; pmid: 6719123
9. C. Scharff, F. Nottebohm, A comparative study of the behavioral deficits following lesions of various parts of the zebra finch song system: Implications for vocal learning. *J. Neurosci.* **11**, 2896–2913 (1991). pmid: 1880555
10. H. Williams, N. Mehta, Changes in adult zebra finch song require a forebrain nucleus that is not necessary for song production. *J. Neurobiol.* **39**, 14–28 (1999). doi: 10.1002/(SICI)1097-4695(199904)39:1<14::AID-NEU2>3.0.CO;2-X; pmid: 10213450
11. M. S. Brainard, A. J. Doupe, Interruption of a basal ganglia-forebrain circuit prevents plasticity of learned vocalizations. *Nature* **404**, 762–766 (2000). doi: 10.1038/35008083; pmid: 10783889
12. B. P. Olveczky, A. S. Andalman, M. S. Fee, Vocal experimentation in the juvenile songbird requires a basal ganglia circuit. *PLOS Biol.* **3**, e153 (2005). doi: 10.1371/journal.pbio.0030153; pmid: 15826219
13. D. Aronov, A. S. Andalman, M. S. Fee, A specialized forebrain circuit for vocal babbling in the juvenile songbird. *Science* **320**, 630–634 (2008). doi: 10.1126/science.1155140; pmid: 18451295
14. S. W. Bottjer, F. Johnson, Circuits, hormones, and learning: Vocal behavior in songbirds. *J. Neurobiol.* **33**, 602–618 (1997). doi: 10.1002/(SICI)1097-4695(19971105)33:5<602::AID-NEU8>3.0.CO;2-8; pmid: 9369462
15. A. R. Pfenning *et al*., Convergent transcriptional specializations in the brains of humans and song learning birds. *Science* **346**, 1256846 (2014).
16. C. Scharff, J. Petri, Evo-devo, deep homology and FoxP2: Implications for the evolution of speech and language. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **366**, 2124–2140 (2011). doi: 10.1098/rstb.2011.0001; pmid: 21690130
17. E. D. Jarvis, F. Nottebohm, Motor-driven gene expression. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 4097–4102 (1997). doi: 10.1073/pnas.94.8.4097; pmid: 9108111
18. R. R. Kimpo, A. J. Doupe, FOS is induced by singing in distinct neuronal populations in a motor network. *Neuron* **18**, 315–325 (1997). doi: 10.1016/S0896-6273(00)80271-8; pmid: 9052801
19. H. Jin, D. F. Clayton, Localized changes in immediate-early gene regulation during sensory and motor learning in zebra finches. *Neuron* **19**, 1049–1059 (1997). doi: 10.1016/S0896-6273(00)80396-7; pmid: 9390518
20. E. D. Jarvis, C. Scharff, M. R. Grossman, J. A. Ramos, F. Nottebohm, For whom the bird sings: Context-dependent gene expression. *Neuron* **21**, 775–788 (1998). doi: 10.1016/S0896-6273(00)80594-2; pmid: 9808464
21. O. Whitney, K. Soderstrom, F. Johnson, Post-transcriptional regulation of zenk expression associated with zebra finch vocal development. *Brain Res. Mol. Brain Res.* **80**, 279–290 (2000). doi: 10.1016/S0169-328X(00)00178-9; pmid: 11038263
22. O. Whitney, F. Johnson, Motor-induced transcription but sensory-regulated translation of ZENK in socially interactive songbirds. *J. Neurobiol.* **65**, 251–259 (2005). doi: 10.1002/neu.20187; pmid: 16155900
23. L. Kubikova, E. A. Turner, E. D. Jarvis, The pallial basal ganglia pathway modulates the behaviorally driven gene expression of the motor pathway. *Eur. J. Neurosci.* **25**, 2145–2160 (2007). doi: 10.1111/j.1460-9568.2007.05368.x; pmid: 17419760
24. K. Wada *et al*., A molecular neuroethological approach for identifying and characterizing a cascade of behaviorally regulated genes. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 15212–15217 (2006). doi: 10.1073/pnas.0607098103; pmid: 17018643
25. X. C. Li, E. D. Jarvis, B. Alvarez-Borda, D. A. Lim, F. Nottebohm, A relationship between behavior, neurotrophin expression, and new neuron survival. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 8584–8589 (2000). doi: 10.1073/pnas.140222497; pmid: 10890902
26. A. J. Lombardino, X.-C. Li, M. Hertel, F. Nottebohm, Replaceable neurons and neurodegenerative disease share depressed UCHL1 levels. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 8036–8041 (2005). doi: 10.1073/pnas.0503239102; pmid: 15911766
27. I. Teramitsu, S. A. White, FoxP2 regulation during undirected singing in adult songbirds. *J. Neurosci.* **26**, 7390–7394 (2006). doi: 10.1523/JNEUROSCI.1662-06.2006; pmid: 16837586
28. J. E. Miller *et al*., Birdsong decreases protein levels of FoxP2, a molecule required for human speech. *J. Neurophysiol.* **100**, 2015–2025 (2008). doi: 10.1152/jn.90415.2008; pmid: 18701760
29. W. C. Warren *et al*., The genome of a songbird. *Nature* **464**, 757–762 (2010). doi: 10.1038/nature08819; pmid: 20360741
30. A. T. Hilliard, J. E. Miller, E. R. Fraley, S. Horvath, S. A. White, Molecular microcircuitry underlies functional specification in a basal ganglia circuit dedicated to vocal learning. *Neuron* **73**, 537–552 (2012). doi: 10.1016/j.neuron.2012.01.005; pmid: 22325205
31. G. Zhang *et al*., Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* **346**, 1311–1320 (2014).
32. S. Haesler *et al*., FoxP2 expression in avian vocal learners and non-learners. *J. Neurosci.* **24**, 3164–3175 (2004). doi: 10.1523/JNEUROSCI.4369-03.2004; pmid: 15056696
33. K. Wada, H. Sakaguchi, E. D. Jarvis, M. Hagiwara, Differential expression of glutamate receptors in avian neural pathways for learned vocalization. *J. Comp. Neurol.* **476**, 44–64 (2004). doi: 10.1002/cne.20201; pmid: 15236466
34. L. Kubikova, K. Wada, E. D. Jarvis, Dopamine receptors in a songbird brain. *J. Comp. Neurol.* **518**, 741–769 (2010). doi: 10.1002/cne.22255; pmid: 20058221
35. X. Li *et al*., Genomic resources for songbird research and their use in characterizing gene expression during brain development. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 6834–6839 (2007). doi: 10.1073/pnas.0701619104; pmid: 17426146
36. P. V. Lovell, D. F. Clayton, K. L. Replogle, C. V. Mello, Birdsong "transcriptomics": Neurochemical specializations of the oscine song system. *PLOS ONE* **3**, e3440 (2008). doi: 10.1371/journal.pone.0003440; pmid: 18941504
37. E. D. Jarvis *et al*., Global view of the functional molecular organization of the avian cerebrum: Mirror images and functional columns. *J. Comp. Neurol.* **521**, 3614–3665 (2013). doi: 10.1002/cne.23404; pmid: 23818122
38. M. R. Lyons, A. E. West, Mechanisms of specificity in neuronal activity-regulated gene transcription. *Prog. Neurobiol.* **94**, 259–295 (2011). doi: 10.1016/j.pneurobio.2011.05.003; pmid: 21620929
39. S. W. Flavell *et al*., Genome-wide analysis of MEF2 transcriptional program reveals synaptic target genes and neuronal activity-dependent polyadenylation site selection. *Neuron* **60**, 1022–1038 (2008). doi: 10.1016/j.neuron.2008.11.029; pmid: 19109909
40. T.-K. Kim *et al*., Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182–187 (2010). doi: 10.1038/nature09033; pmid: 20393465
41. S. Dong *et al*., Discrete molecular states in the brain accompany changing responses to a vocal signal. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 11364–11369 (2009). doi: 10.1073/pnas.0812998106; pmid: 19541599
42. G. Feenders *et al*., Molecular mapping of movement-associated areas in the avian brain: A motor theory for vocal learning origin. *PLOS ONE* **3**, e1768 (2008). doi: 10.1371/journal.pone.0001768; pmid: 18335043
43. C. S. Hill, R. Treisman, Growth factors and gene expression: Fresh insights from arrays. *Sci. STKE* **1999**, PE1 (1999). doi: 10.1126/stke.1999.3.pe1; pmid: 11865182
44. A. Barco, J. M. Alarcon, E. R. Kandel, Expression of constitutively active CREB protein facilitates the late phase of long-term potentiation by enhancing synaptic capture. *Cell* **108**, 689–703 (2002). doi: 10.1016/S0092-8674(02)00657-8; pmid: 11893339
45. S. Sinha, On counting position weight matrix matches in a sequence, with application to discriminative motif finding. *Bioinformatics* **22**, e454–e463 (2006). doi: 10.1093/bioinformatics/btl227; pmid: 16873507
46. J. Kim *et al*., Functional characterization of transcription factor motifs using cross-species comparison across large evolutionary distances. *PLOS Comput. Biol.* **6**, e1000652 (2010). doi: 10.1371/journal.pcbi.1000652; pmid: 20126523
47. A. R. Pfenning, R. Schwartz, A. L. Barth, A comparative genomics approach to identifying the plasticity transcriptome. *BMC Neurosci.* **8**, 20 (2007). doi: 10.1186/1471-2202-8-20; pmid: 17355637
48. A. R. Pfenning *et al*., Genome-wide identification of calcium-response factor (CaRF) binding sites predicts a role in regulation of neuronal signaling pathways. *PLOS ONE* **5**, e10870 (2010). doi: 10.1371/journal.pone.0010870; pmid: 20523734
49. S. W. Flavell, M. E. Greenberg, Signaling mechanisms linking neuronal activity to gene expression and plasticity of the nervous system. *Annu. Rev. Neurosci.* **31**, 563–590 (2008). doi: 10.1146/annurev.neuro.31.060407.125631; pmid: 18558867
50. S. Loebrich, E. Nedivi, The function of activity-regulated genes in the nervous system. *Physiol. Rev.* **89**, 1079–1103 (2009). doi: 10.1152/physrev.00013.2009; pmid: 19789377
51. X. Cao, R. Mahendran, G. R. Guy, Y. H. Tan, Detection and characterization of cellular EGR-1 binding to its recognition site. *J. Biol. Chem.* **268**, 16949–16957 (1993). pmid: 8349585
52. O. Johnsen *et al*., Small Maf proteins interact with the human transcription factor TCF11/Nrf1/LCR-F1. *Nucleic Acids Res.* **24**, 4289–4297 (1996). doi: 10.1093/nar/24.21.4289; pmid: 8932385
53. A. Subramanian *et al*., Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15545–15550 (2005). doi: 10.1073/pnas.0506580102; pmid: 16199517
54. K. A. McDowell *et al*., Reduced cortical BDNF expression and aberrant memory in Carf knock-out mice. *J. Neurosci.* **30**, 7453–7465 (2010). doi: 10.1523/JNEUROSCI.3997-09.2010; pmid: 20519520
55. ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012). doi: 10.1038/nature11247; pmid: 22955616
56. UCSC Genome Browser; https://genome.ucsc.edu/cgi-bin/hgTracks?hgS_doOtherUser=submit&hgS_otherUserName=apfenning&hgS_otherUserSessionName=taeGut1_basic (2014).

57. P. Goelet, V. F. Castellucci, S. Schacher, E. R. Kandel, The long and the short of long-term memory—A molecular framework. *Nature* **322**, 419–422 (1986). doi: 10.1038/322419a0; pmid: 2874497
58. A. N. Malik *et al*., Genome-wide identification and characterization of functional neuronal activity-dependent enhancers. *Nat. Neurosci.* **17**, 1330–1339 (2014). doi: 10.1038/nn.3808; pmid: 25195102
59. A. E. West, M. E. Greenberg, Neuronal activity-regulated gene transcription in synapse development and cognitive function. *Cold Spring Harb. Perspect. Biol.* **3**, a005744–a005744 (2011). doi: 10.1101/cshperspect.a005744; pmid: 21555405
60. I. Spiegel *et al*., Npas4 regulates excitatory-inhibitory balance within neural circuits through cell-type-specific gene programs. *Cell* **157**, 1216–1229 (2014). doi: 10.1016/j.cell.2014.03.058; pmid: 24855953
61. H. Horita, K. Wada, M. V. Rivas, E. Hara, E. D. Jarvis, The dusp1 immediate early gene is regulated by natural stimuli predominantly in sensory input neurons. *J. Comp. Neurol.* **518**, 2873–2901 (2010). pmid: 20506480
62. H. Horita *et al*., Specialized motor-driven dusp1 expression in the song systems of multiple lineages of vocal learning birds. *PLOS ONE* **7**, e42173 (2012). doi: 10.1371/journal.pone.0042173; pmid: 22876306
63. K. Sakamoto, K. Karelina, K. Obrietan, CREB: A multifaceted regulator of neuronal plasticity and protection. *J. Neurochem.* **116**, 1–9 (2011). doi: 10.1111/j.1471-4159.2010.07080.x; pmid: 21044077
64. J. I. Morgan, T. Curran, Stimulus-transcription coupling in neurons: Role of cellular immediate-early genes. *Trends Neurosci.* **12**, 459–462 (1989). doi: 10.1016/0166-2236(89)90096-9; pmid: 2479148
65. C. H. Bailey, D. Bartsch, E. R. Kandel, Toward a molecular definition of long-term memory storage. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 13445–13452 (1996). doi: 10.1073/pnas.93.24.13445; pmid: 8942955
66. J. M. Wettenhall, G. K. Smyth, limmaGUI: A graphical user interface for linear modeling of microarray data. *Bioinformatics* **20**, 3705–3706 (2004). doi: 10.1093/bioinformatics/bth449; pmid: 15297296

ACKNOWLEDGMENTS

We thank S. Augustin, J. Coleman, and N. Diaz Nelson for help collecting and processing singing bird samples for microarray analysis; H. Dressman and L.-L. Rowlette at the Duke Institute for Genome Science and Policy for microarray hybridizations; We thank A. Cavanaugh, M. Lawson, and M. Dean for useful comments after careful reading of an earlier version of this manuscript. O.W. was supported by postdoctoral training grants from American Psychological Association fellowship funded by the National Institute of Mental Health grant 5T32MH018882-18 and the National Science Foundation grant 0610337. This work was supported by grants from the National Institute on Deafness and Other Communication Disorders grant R01DC007218 and the Howard Hughes Medical Institute to E.D.J.. Microarray data have been submitted to the Gene Expression Omnibus (www.ncbi.nlm.nih.gov/geo) under accession no. GSE33365. Annotations are also available in table S1. Raw files of the Chip-seq H3K27ac experiments are available at www.broadinstitute.org/~pfenning/finchWig.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/346/6215/1256780/suppl/DC1
Materials and Methods
Figs. S1 to S12
Tables S1 to S22
References (*67*–*117*)

2 June 2014; accepted 11 November 2014
10.1126/science.1256780

RESEARCH ARTICLE SUMMARY

AVIAN GENOMICS

Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs

Richard E. Green,* Edward L. Braun, Joel Armstrong, Dent Earl, Ngan Nguyen, Glenn Hickey, Michael W. Vandewege, John A. St. John, Salvador Capella-Gutiérrez, Todd A. Castoe, Colin Kern, Matthew K. Fujita, Juan C. Opazo, Jerzy Jurka, Kenji K. Kojima, Juan Caballero, Robert M. Hubley, Arian F. Smit, Roy N. Platt, Christine A. Lavoie, Meganathan P. Ramakodi, John W. Finger Jr., Alexander Suh, Sally R. Isberg, Lee Miles, Amanda Y. Chong, Weerachai Jaratlerdsiri, Jaime Gongora, Christopher Moran, Andrés Iriarte, John McCormack, Shane C. Burgess, Scott V. Edwards, Eric Lyons, Christina Williams, Matthew Breen, Jason T. Howard, Cathy R. Gresham, Daniel G. Peterson, Jürgen Schmitz, David D. Pollock, David Haussler, Eric W. Triplett, Guojie Zhang, Naoki Irie, Erich D. Jarvis, Christopher A. Brochu, Carl J. Schmidt, Fiona M. McCarthy, Brant C. Faircloth, Federico G. Hoffmann, Travis C. Glenn, Toni Gabaldón, Benedict Paten, David A. Ray*

INTRODUCTION: Crocodilians and birds are the two extant clades of archosaurs, a group that includes the extinct dinosaurs and pterosaurs. Fossils suggest that living crocodilians (alligators, crocodiles, and gharials) have a most recent common ancestor 80 to 100 million years ago. Extant crocodilians are notable for their distinct morphology, limited intraspecific variation, and slow karyotype evolution. Despite their unique biology and phylogenetic position, little is known about genome evolution within crocodilians.

ON OUR WEB SITE

Read the full article at http://dx.doi.org/10.1126/science.1254449

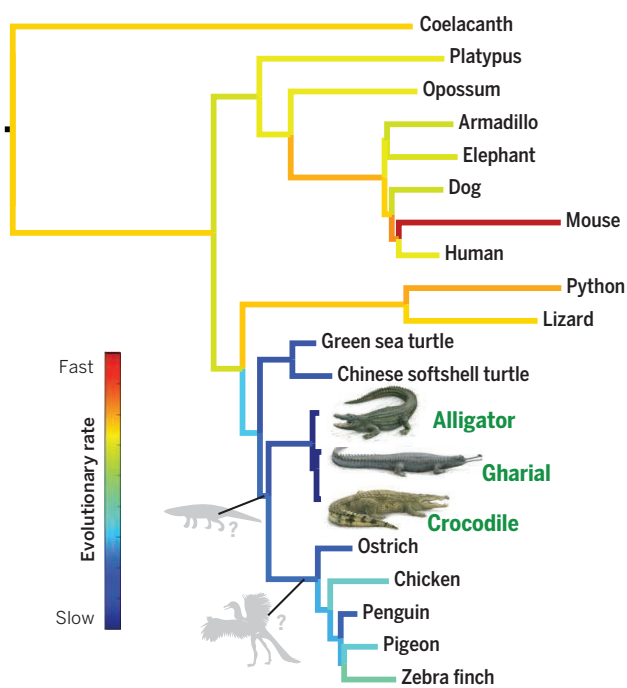
RATIONALE: Genome sequences for the American alligator, saltwater crocodile, and Indian gharial—representatives of all three extant crocodilian families—were obtained to facilitate better understanding of the unique biology of this group and provide a context for studying avian genome evolution. Sequence data from these three crocodilians and birds also allow reconstruction of the ancestral archosaurian genome.

RESULTS: We sequenced shotgun genomic libraries from each species and used a variety of assembly strategies to obtain draft genomes for these three crocodilians. The assembled scaffold N50 was highest for the alligator (508 kilobases). Using a panel of reptile genome sequences, we generated phylogenies that confirm the sister relationship between crocodiles and gharials, the relationship with birds as members of extant Archosauria, and the outgroup status of turtles relative to birds and crocodilians.

We also estimated evolutionary rates along branches of the tetrapod phylogeny using two approaches: ultraconserved element–anchored sequences and fourfold degenerate sites within stringently filtered orthologous gene alignments. Both analyses indicate that the rates of base substitution along the crocodilian and turtle lineages are extremely low. Supporting observations were made for transposable element content and for gene family evolution. Analysis of whole-genome alignments across a panel of reptiles and mammals showed that the rate of accumulation of microinsertions and microdeletions is proportionally lower in crocodilians, consistent with a single underlying cause of a reduced rate of evolutionary change rather than intrinsic differences in base-repair machinery. We hypothesize that this single cause may be a consistently longer generation time over the evolutionary history of Crocodylia.

Low heterozygosity was observed in each genome, consistent with previous analyses, including the Chinese alligator. Pairwise sequential Markov chain analysis of regional heterozygosity indicates that during glacial cycles of the Pleistocene, each species suffered reductions in effective population size. The reduction was especially strong for the American alligator, whose current range extends farthest into regions of temperate climates.

CONCLUSION: We used crocodilian, avian, and outgroup genomes to reconstruct 584 megabases of the archosaurian common ancestor genome and the genomes of key ancestral nodes. The estimated accuracy of the archosaurian genome reconstruction is 91% and is higher for conserved regions such as genes. The reconstructed genome can be improved by adding more crocodilian and avian genome assemblies and may provide a unique window to the genomes of extinct organisms such as dinosaurs and pterosaurs. ■



Evolutionary rates of tetrapods inferred from DNA sequences anchored by ultraconserved elements. Evolutionary rates among reptiles vary, with especially low rates among extant crocodilians but high rates among squamates. We have reconstructed the genomes of the common ancestor of birds and of all archosaurs (shown in gray silhouette, although the morphology of these species is uncertain).

A complete list of affiliations is available in the full article online.
*Corresponding author. E-mail: ed@soe.ucsc.edu (R.E.G.); david.a.ray@ttu.edu (D.A.R.)
Cite this article as R. E. Green *et al.*, *Science* **346**, 1254449 (2014). DOI: 10.1126/science.1254449

Downloaded from www.sciencemag.org on December 14, 2014

Published by AAAS

RESEARCH ARTICLE

Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs

Richard E. Green,[1]* Edward L. Braun,[2] Joel Armstrong,[1,3] Dent Earl,[1,3] Ngan Nguyen,[1,3] Glenn Hickey,[1,3] Michael W. Vandewege,[4] John A. St. John,[1]§ Salvador Capella-Gutiérrez,[5,6] Todd A. Castoe,[7,8] Colin Kern,[9] Matthew K. Fujita,[8] Juan C. Opazo,[10] Jerzy Jurka,[11]† Kenji K. Kojima,[11] Juan Caballero,[12] Robert M. Hubley,[12] Arian F. Smit,[12] Roy N. Platt,[4,13] Christine A. Lavoie,[4] Meganathan P. Ramakodi,[4,13]‡ John W. Finger Jr.,[14] Alexander Suh,[15,16] Sally R. Isberg,[17,18,19] Lee Miles,[18]# Amanda Y. Chong,[18] Weerachai Jaratlerdsiri,[18] Jaime Gongora,[18] Christopher Moran,[18] Andrés Iriarte,[20] John McCormack,[21] Shane C. Burgess,[22] Scott V. Edwards,[23] Eric Lyons,[24] Christina Williams,[25] Matthew Breen,[25] Jason T. Howard,[26] Cathy R. Gresham,[13] Daniel G. Peterson,[13,27] Jürgen Schmitz,[15] David D. Pollock,[7] David Haussler,[3,28] Eric W. Triplett,[29] Guojie Zhang,[30,31] Naoki Irie,[32] Erich D. Jarvis,[26] Christopher A. Brochu,[33] Carl J. Schmidt,[34] Fiona M. McCarthy,[35] Brant C. Faircloth,[36,37] Federico G. Hoffmann,[4,13] Travis C. Glenn,[14] Toni Gabaldón,[5,6,38] Benedict Paten,[3] David A. Ray[4,13,39]*

To provide context for the diversification of archosaurs—the group that includes crocodilians, dinosaurs, and birds—we generated draft genomes of three crocodilians: *Alligator mississippiensis* (the American alligator), *Crocodylus porosus* (the saltwater crocodile), and *Gavialis gangeticus* (the Indian gharial). We observed an exceptionally slow rate of genome evolution within crocodilians at all levels, including nucleotide substitutions, indels, transposable element content and movement, gene family evolution, and chromosomal synteny. When placed within the context of related taxa including birds and turtles, this suggests that the common ancestor of all of these taxa also exhibited slow genome evolution and that the comparatively rapid evolution is derived in birds. The data also provided the opportunity to analyze heterozygosity in crocodilians, which indicates a likely reduction in population size for all three taxa through the Pleistocene. Finally, these data combined with newly published bird genomes allowed us to reconstruct the partial genome of the common ancestor of archosaurs, thereby providing a tool to investigate the genetic starting material of crocodilians, birds, and dinosaurs.

Crocodilians, birds, dinosaurs, and pterosaurs are a monophyletic group known as the archosaurs. Crocodilians and birds are the only extant members; thus, crocodilians (alligators, caimans, crocodiles, and gharials) are the closest living relatives of all birds (*1*, *2*). Although crocodilians diverged from birds more than 240 million years ago (Ma), animals with morphology unambiguously similar to the extant crocodilian families (Alligatoridae, Crocodylidae, and Gavialidae) first appear in the fossil record between 80 and 90 Ma (*3*). Unlike other vertebrates such as mammals, squamates, and birds, which underwent substantial diversification, extant crocodilian species have maintained morphological and ecological similarities (*4*). Slow divergence among living crocodilians is also observed at the level of karyotype evolution (*5*).

Crocodilians are important model organisms in fields as diverse as developmental biology, osmoregulation, cardiophysiology, paleoclimatology, sex determination, population genetics, paleobiogeography, and functional morphology (*4*). For example, the males and females of all crocodilians (like some but not all reptiles) are genetically identical; sexual fate is determined during development by a temperature-sensing mechanism whose molecular basis remains poorly understood (*6*). More broadly, reptilian genomes exhibit substantial variation in isochore content, chromosome sizes and compositions (e.g., some but not all species have GC-rich and gene-rich microchromosomes), and sex determination mechanisms. Remarkably, this plasticity in large-scale genome features is often coincident with a slower rate of karyotype and sequence evolution (*7*).

We sequenced the genomes of the American alligator, the saltwater crocodile, and the Indian gharial, spanning the three major extant crocodilian lineages (*3*, *8*–*10*). These crocodilian genomes augment the list of assembled genomes from avian and nonavian reptiles (*11*–*16*), allowing us to probe the lineage-specific novelties in avian and crocodilian evolution. They also provide the substrate for computational inference of the common ancestor archosaur genome.

Genome assembly and annotation

We generated high-coverage Illumina sequence data (tables S1 to S3) from paired-end and mate-pair libraries from each species: alligator, crocodile, and gharial. The assembly strategy for each taxon differed because of varying legacy data and developments in library preparation methods during the course of the project (*17*). Genome scaffolding of the alligator (and, to a lesser extent, the crocodile) was aided by the availability

[1]Department of Biomolecular Engineering, University of California, Santa Cruz, CA 95064, USA. [2]Department of Biology and Genetics Institute, University of Florida, Gainesville, FL 32611, USA. [3]Center for Biomolecular Science and Engineering, University of California, Santa Cruz, CA 95064, USA. [4]Department of Biochemistry, Molecular Biology, Entomology and Plant Pathology, Mississippi State University, Mississippi State, MS 39762, USA. [5]Bioinformatics and Genomics Programme, Centre for Genomic Regulation, 08003 Barcelona, Spain. [6]Universitat Pompeu Fabra, 08003 Barcelona, Spain. [7]Department of Biochemistry and Molecular Genetics, University of Colorado School of Medicine, Aurora, CO 80045, USA. [8]Department of Biology, University of Texas, Arlington, TX 76019, USA. [9]Department of Computer and Information Sciences, University of Delaware, Newark, DE 19717, USA. [10]Instituto de Ciencias Ambientales y Evolutivas, Facultad de Ciencias, Universidad Austral de Chile, Valdivia, Chile. [11]Genetic Information Research Institute, Mountain View, CA 94043, USA. [12]Institute for Systems Biology, Seattle, WA 98109, USA. [13]Institute for Genomics, Biocomputing and Biotechnology, Mississippi State University, Mississippi State, MS 39762, USA. [14]Department of Environmental Health Science, University of Georgia, Athens, GA 30602, USA. [15]Institute of Experimental Pathology (ZMBE), University of Münster, D-48149 Münster, Germany. [16]Department of Evolutionary Biology (EBC), Uppsala University, SE-752 36 Uppsala, Sweden. [17]Porosus Pty. Ltd., Palmerston, NT 0831, Australia. [18]Faculty of Veterinary Science, University of Sydney, Sydney, NSW 2006, Australia. [19]Centre for Crocodile Research, Noonamah, NT 0837, Australia. [20]Departamento de Desarrollo Biotecnológico, Instituto de Higiene, Facultad de Medicina, Universidad de la República, Montevideo, Uruguay. [21]Moore Laboratory of Zoology, Occidental College, Los Angeles, CA 90041, USA. [22]College of Agriculture and Life Sciences, University of Arizona, Tucson, AZ 85721, USA. [23]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA. [24]School of Plant Sciences, University of Arizona, Tucson, AZ 85721, USA. [25]Department of Molecular Biomedical Sciences, North Carolina State University, Raleigh, NC 27607, USA. [26]Howard Hughes Medical Institute, Department of Neurobiology, Duke University Medical Center, Durham, NC 27710, USA. [27]Department of Plant and Soil Sciences, Mississippi State University, Mississippi State, MS 39762, USA. [28]Howard Hughes Medical Institute, Bethesda, MD 20814, USA. [29]Department of Microbiology and Cell Science, University of Florida, Gainesville, FL 32611, USA. [30]China National GeneBank, BGI-Shenzhen, Shenzhen, China. [31]Center for Social Evolution, Department of Biology, University of Copenhagen, Copenhagen, Denmark. [32]Department of Biological Sciences, Graduate School of Science, University of Tokyo, Tokyo, Japan. [33]Department of Earth and Environmental Sciences, University of Iowa, Iowa City, IA 52242, USA. [34]Department of Animal and Food Sciences, University of Delaware, Newark, DE 19717, USA. [35]School of Animal and Comparative Biomedical Sciences, University of Arizona, Tucson, AZ 85721, USA. [36]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90019, USA. [37]Department of Biological Sciences, Louisiana State University, Baton Rouge, LA 70803, USA. [38]Institució Catalana de Recerca i Estudis Avançats, 08010 Barcelona, Spain. [39]Department of Biological Sciences, Texas Tech University, Lubbock, TX 79409, USA.

***Corresponding author. E-mail: ed@soe.ucsc.edu (R.E.G.); david.a.ray@ttu.edu (D.A.R.)** †Deceased. ‡Present address: Cancer Prevention and Control, Fox Chase Cancer Center, Philadelphia, PA 19111, USA, and Department of Biology, Temple University, Philadelphia, PA 19122, USA. §Present address: Driver Group LLC, 1700 Owens Street, Suite 200, San Francisco, CA 94158, USA. #Present address: FB Rice Patent & Trade Mark Attorneys, 44 Market Street, Sydney, NSW 2000, Australia.

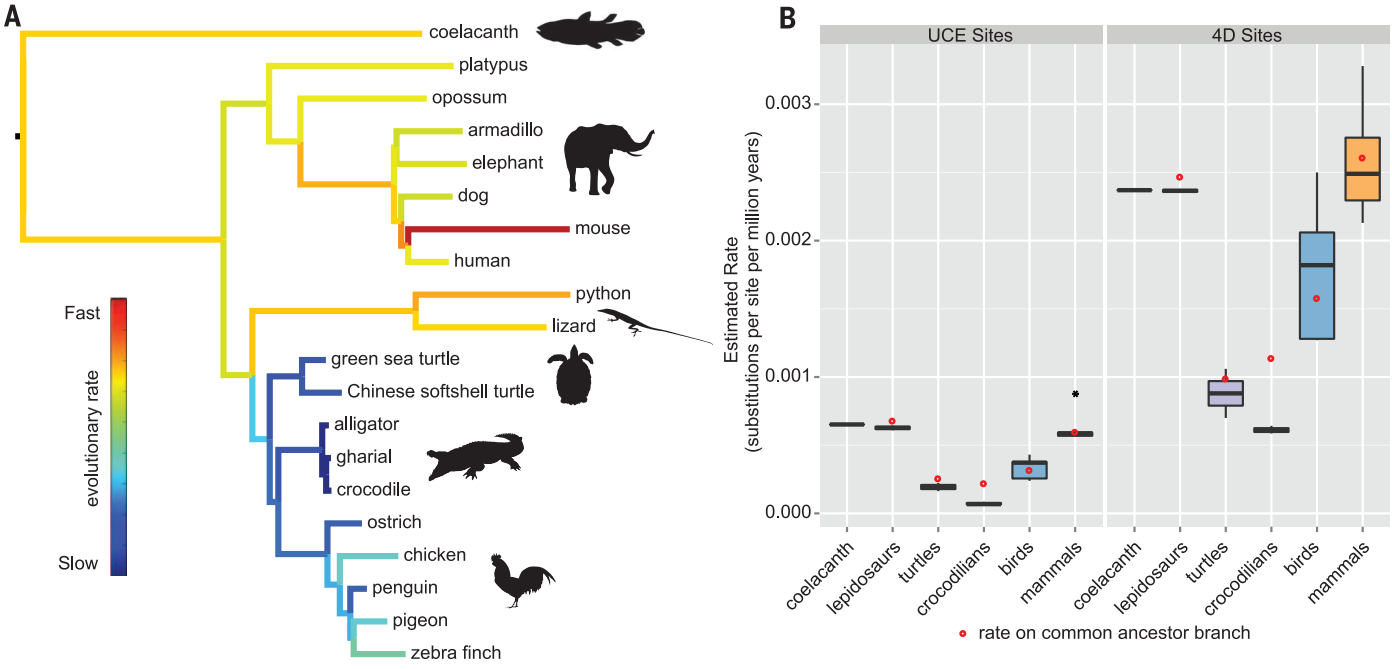


Fig. 1. Rates of substitution for ultraconserved elements (UCEs) and fourfold degenerate (4D) sites. (**A**) Inferred amniote phylogeny based on maximum likelihood analysis of partitioned UCE-anchored loci using RAxML v7.3.4 (*17*). All branches received 100% bootstrap support. Colors indicate the estimated rates, with cooler colors corresponding to lower rates of molecular evolution. (**B**) Estimated rates of molecular evolution for UCE-anchored loci (left) and 4D sites (right). Red dots indicate the estimated rate for the branch ancestral to the group of interest. The UCE rate for mouse is an outlier and is indicated by a black dot.

of bacterial artificial chromosome (BAC) sequences and BAC end-sequence data. RNA-seq data were collected from the alligator and, to a lesser extent, the crocodile and gharial (*17*). Stringently filtered consensus gene sequences were used for quality assessment of drafts of the genome assemblies and finally to aid in scaffolding the assemblies. Details of the libraries and assembly statistics for each genome are summarized in tables S1 to S4.

Gene annotation was accomplished using a combination of transcriptome sequencing (RNA-seq) data and homology-based analyses (*17*). We identified 23,323 protein-coding genes in the alligator, 13,321 in the crocodile, and 14,043 in the gharial (table S5). The unevenness likely reflects the larger overall scaffold size (N50) of the alligator genome assembly (table S4) and the predominance of alligator transcriptome data used to guide gene identification (table S6). This unevenness of annotation complicates direct comparisons of gene content. Therefore, for protein-coding sequence analyses, we compared orthologous sequence of the crocodile and gharial to the more thoroughly annotated alligator genome. We assigned names to 55% of crocodilian genes on the basis of orthology to vertebrates with existing standardized nomenclature (human, mouse, anole, chicken, and zebrafish). Between 60 and 70% of crocodilian proteins had conserved functional motifs on the basis of comparison to other vertebrates, and we provided 377,441 Gene Ontology (GO) annotations for 43,436 crocodilian proteins.

Transposable elements (TEs) were identified de novo in all three crocodilians, and analyses resulted in a library of 1269 different TEs (table S7)—a large number for a vertebrate. This high TE count in crocodilian genomes is attributable, at least in part, to the apparently low rate of base substitution in crocodilians, as discussed below. We find that ~37.5% of each crocodilian genome can be annotated as TEs (table S7), a value intermediate between mammals (40 to 60%) and birds (12 to 15%) (*18*–*23*).

Ultraconserved element phylogeny and molecular evolution

Ultraconserved elements (UCEs) were originally defined as orthologous segments that exhibit very high levels of sequence conservation (*24*). Subsequent work established that UCEs often occur in single-copy regions of the genome. Regions immediately flanking the core of a UCE typically exhibit progressively greater evolutionary rates (*25*–*27*). The relative ease of assessing orthology for UCEs and their flanking regions (hereafter called UCE-anchored loci), combined with their ease of alignment and the fact that they exhibit little or no substitution saturation, makes them useful for estimating relative evolutionary rates across all tetrapods. We identified and extracted 965 UCE-anchored sequences from the three crocodilian genomes and compared them to their orthologs from representatives of all major tetrapod lineages [in addition to the archosaurs, we included mammals, lepidosaurs (lizards and snakes), turtles, and an amphibian along with the coelacanth outgroup (*17*) (table S8)]. Using these data, we inferred tetrapod phylogeny and examined rates of evolution along the branches (Fig. 1A and figs. S1 to S7).

The phylogeny estimated using UCE-anchored data largely agrees with other studies (*8*, *10*, *28*, *29*). For example, we recovered Longirostres (crocodiles + gharials) within Crocodylia, found crocodilians to be the sister group of birds (supporting the clade Archosauria), and confirmed turtles as the sister group to living archosaurs. Branch lengths across this phylogeny suggest that crocodilians exhibit a low rate of molecular evolution for UCE-anchored loci relative to all tetrapod groups (Fig. 1A), including the slowly evolving turtles. To explore the evolutionary tempo of crocodilians, we used divergence time estimates for critical nodes to derive estimates of absolute substitution rates across the tree (*17*). These estimates suggest that the molecular evolution of crocodilians is slower than that of all other lineages (figs. S2, S4, S7, S15, and S16). Indeed, the crocodilian rate is approximately an order of magnitude slower than that of lepidosaurs and mammals. Perhaps more important, the availability of multiple bird, crocodilian, and turtle genomes allows us to estimate the ancestral rates for these groups (Fig. 1B). Using a variety of calibration times for the TMRCA (time to most recent common ancestor) of birds, crocodilians, and archosaurs (fig. S8 and table S9), we find that the rate of UCE evolution for the avian stem lineage was similar to that of extant avian lineages (Fig. 1B and fig. S7). In contrast, the crocodilian stem lineage evolved more rapidly than its extant lineages. Given the low rates observed in both turtles and crocodilians and the reduced rate in the avian stem lineage, we propose that the ancestor of all archosaurs was likely characterized by an extremely slow rate of molecular

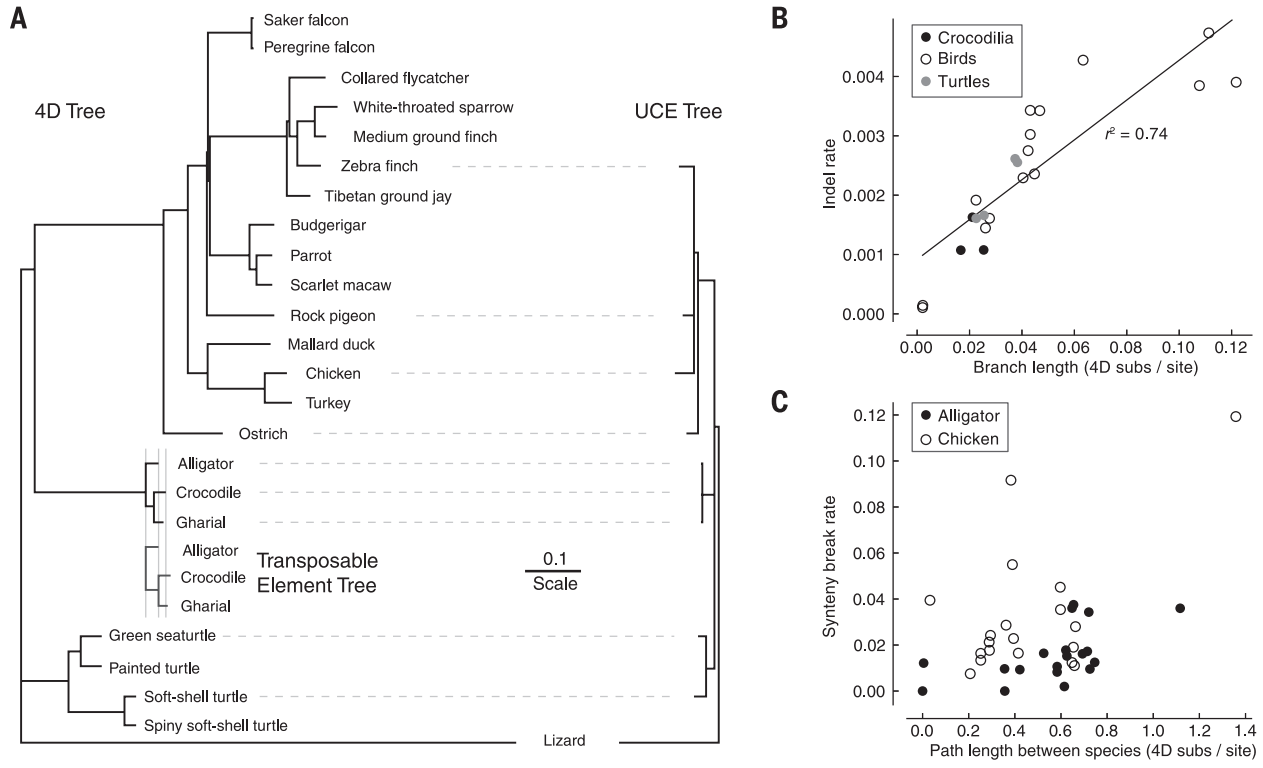


Fig. 2. Rates of substitution, micro-indels, and break-point evolution. (**A**) Rates of substitution at 4D sites, transposable elements (TEs), and, for comparison, UCE-anchored loci. Scale bar denotes substitutions per site. (**B**) Indel rate versus 4D substitutions per site for each extant lineage. (**C**) Gene synteny breakage rate versus 4D substitutions per site, each measured with respect to either alligator or chicken.

evolution that subsequently increased on the avian stem lineage.

Gene-based phylogeny and molecular evolution

We used the PhylomeDB pipeline (*19*) to identify 337 single-copy orthologous gene sequences (*17*) from 22 tetrapod genomes (table S10). Phylogenetic analysis of a concatenated alignment of these genes (fig. S10) produced a tree congruent with the UCE-based phylogeny shown in Fig. 1A and other amniote phylogenies (*28*, *30*). The concatenated alignment of orthologs was then further filtered to extract fourfold degenerate (4D) sites (*17*), which evolve at a rate similar to the neutral rate. Although some 4D sites may be subject to purifying selection (*31*), studies in birds suggest that substitutions at 4D sites accumulate ~75% as rapidly as those at other sites thought to be neutral (*32*). Thus, their rate is expected to be much closer to the neutral rate than the rate estimated using UCE-anchored data. As expected, substitution rates at 4D sites were higher than the rate estimated using UCE-anchored regions (Fig. 1B). However, the pattern of relative rates for different taxa was qualitatively similar to that reconstructed using the UCE-anchored regions (Fig. 1B and figs. S13 to S16).

A larger survey of aligned genes (without the single-copy orthology filters) found 9574 trees that suggested monophyly of birds, turtles, and crocodilians relative to squamates; the vast majority of those (6880; 72%) placed crocodilians and birds together in a clade. Only 28% of trees supported alternate topologies [birds + turtles or crocodilians + turtles (*17*)]. Although the placement of gharial within the crocodilian phylogeny has been contentious over the past several decades (*33*), a clear majority (78.4%) of protein-coding gene trees supported Longirostres (*8*).

Rates of genome evolution in crocodilians, birds, and other reptiles

To explore patterns of molecular evolution across the genomes of crocodilians, we created a whole-genome alignment (WGA) (*17*) that included 23 reptile genomes, including the three crocodilians, 15 birds, four turtles, and the Carolina anole lizard as the outgroup (table S12). Consistent with our other results, the WGA analysis revealed low genome-wide pairwise divergences among crocodilians (table S13); for example, the alligator and crocodile (which shared a common ancestor ~80 to 100 Ma; node O in table S9) have ~93% genome-wide identity. This is similar to the level of identity between human and rhesus macaque, whose common ancestor lived only ~23 Ma (*34*), indicating exceptionally low rates of evolution relative to mammals.

This WGA for birds and reptiles also provides an opportunity to assess the relative rates of different substitution types using a single alignment framework. We compared rates at 4D sites (Fig. 2 and table S14) with those occurring within orthologous TE insertions that are shared among the three crocodilians (table S15). Substitutions in TEs, which presumably accumulate at close to the neutral rate (*35*), accumulated slightly more rapidly than those at 4D sites extracted from the WGA (Fig. 2A). The WGA also allowed us to estimate the rate of micro-indels [≤10 base pairs (bp) per event, filtered to avoid alignment errors] relative to substitutions. This ratio for crocodilians (0.064 micro-indels per substitution) is similar to that in birds and turtles (Fig. 2B) and is within the range of previous estimates for mammals (*36*, *37*). The ratio of microdeletions to micro-insertions was similar across the tree (average ~1.94; table S16) and concordant with previous estimates from other taxa (*36*, *37*), with no apparent bias toward either category in crocodilians, birds, or turtles (table S20).

Finally, we used the multiple-species WGA to examine the conservation of synteny between adjacent gene pairs in chicken and alligator, examining only those pairs where both genes were unambiguously located (*17*). We found high levels of gene order conservation between crocodilian genomes, similar to that between comparably separated bird genomes—a group marked by its extreme syntenic conservation relative to mammals (Fig. 2C). Thus, the low evolutionary rates observed in crocodilians are not specific to substitutions, but also include micro-indels and gene-level rearrangements.

Transposable elements evolve slowly in crocodilians

Of the annotated TEs, 95% belong to families that appear in all three genomes with near-equal frequency. Thus, only ~5% of TE copies (representing <2% of the genome) arose after the split of

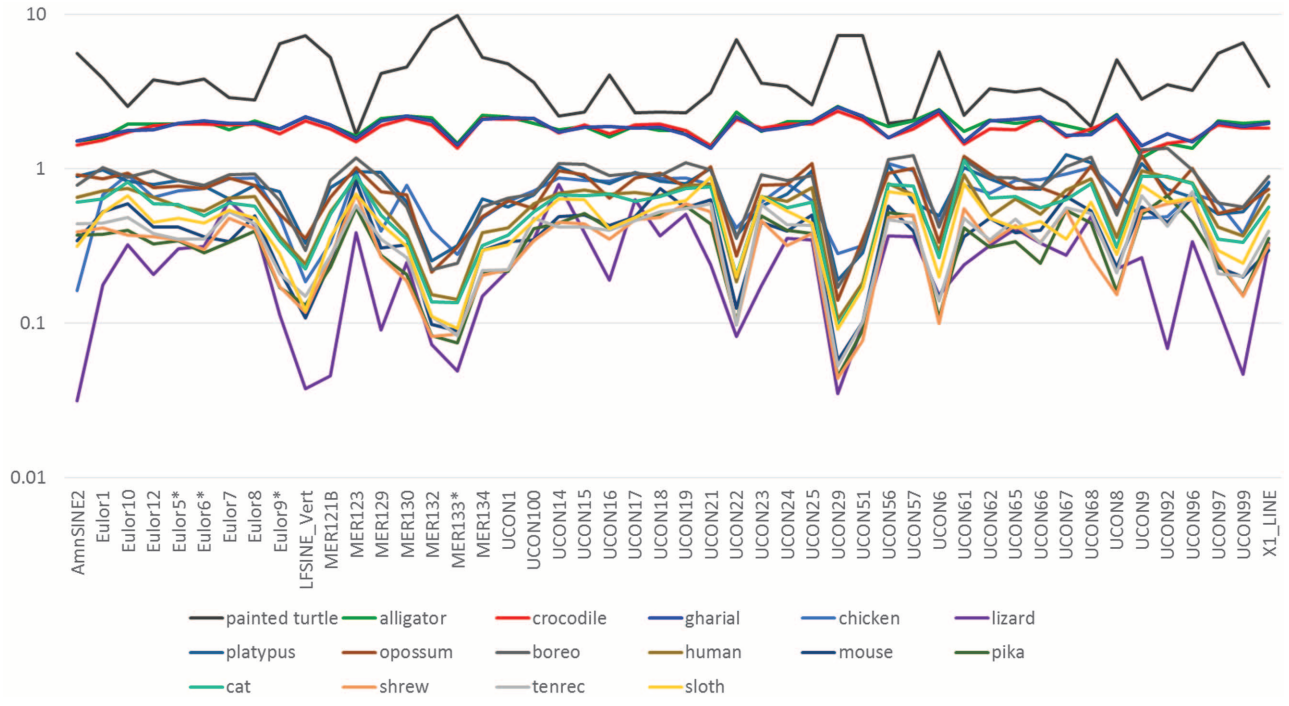


Fig. 3. Relative TE numbers among amniotes. Shown are TE copies that predate the speciation of crocodilians and mammals in 16 amniote genomes. The figure displays 55 unrelated TE families present in all amniote genomes. The numbers of bases, on a log scale, identified in each individual genome relative to the average in all 16 genomes are identified. An asterisk indicates that two or more subfamilies were combined to form a single category. See (*17*) for the full analysis encompassing all 74 TE families.

Longirostres (crocodile + gharial) from alligators, approximately 80 to 100 Ma (table S9). Given that there is an ascertainment bias against older repeats, these data suggest that the rate of new TE family invasion/evolution has generally been decreasing in crocodilians, with the exception of a minor burst of novel activity in the common ancestor of Longirostres (fig. S20). Indeed, in the ~235 million years between the mammal-crocodilian divergence and the origin of crown crocodilians, at least 823 TE families were active—a rate of around 3.5 TE families per million years. The rate has fallen below 1.0 in both crocodile and gharial since their divergence.

The "visibility" of TE copies introduced before the divergence of mammals and reptiles at ~310 to 330 Ma (table S9) (*17*) provides another line of evidence for the extraordinarily low rate of crocodilian genome evolution. Averaged over 74 unrelated families of such elements (*17*), crocodilian genomes contain a considerable amount of DNA that is recognizably derived from TEs: five times the amount in the typical mammalian genome, three times the amount in the reconstructed boreoeutherian (the mammalian clade comprising primates, rodents, carnivores, bats, and a number of additional orders; "boreo" in Fig. 3) genome (*36*), 3.8 times the identifiable amount in the chicken genome, and 15 times that in the anole genome. Surprisingly, relative to crocodilians, the painted turtle genome contains on average 2.3 times as many bases recognizably derived from each of these repeats (Fig. 3 and figs. S21 to S23), suggesting an even slower neutral decay rate. The consistency of the relative representation of these unrelated elements in each genome suggests that these ratios are not the result of differential lineage-specific accumulation but instead represent actual differences in mutation and deletion rates, and that crocodilians exhibit a neutral mutation rate that is among the slowest found in vertebrates and may be the slowest within amniotes.

Gene family evolution suggests retention of ancestral orthologs in crocodilian lineages

We used gene trees from the phylome analysis to search for gene families that underwent duplications within the crocodilian lineage (*17*). Olfactory receptors (ORs) constitute one of the largest vertebrate gene families; they are small, single-exon genes, making them relatively easy to investigate. ORs have also played a central role in the development of our understanding of how gene families evolve (*38*). Similar to results found in other amniote lineages (*39*, *40*), genes associated with olfactory perception were overrepresented among duplicated genes in crocodilians. Crocodilians possess a diverse OR repertoire, and each species has about 1000 ORs, half of which are likely functional (table S21) (*17*); this is not unusual for a tetrapod genome. However, in other tetrapods the ORs derive from independent expansions of a small number of ancestral OR genes within those lineages (*38*), as we observed for the birds and turtles we examined (Fig. 4 and fig. S24). In contrast, crocodilian OR repertoires almost exclusively reflect the retention of OR genes present in the common ancestor of the crown crocodilian, followed by a few gains or losses (Fig. 4 and fig. S24). This observation—many retained ancestral genes rather than independent expansion—suggests that crocodilians have achieved a diverse OR repertoire using a strategy of retention of ancient genes, as opposed to the generation of novel variants.

Genetic diversity and natural history of Crocodylia

We used the genomic data generated here to investigate the population history of each crocodilian species. Mapping shotgun reads back to the assembly, we identified and quantified the rate of heterozygosity (*17*) within each species. All three genomes exhibited a low degree of heterozygosity relative to most mammalian and avian genomes (Fig. 5). Among the three crocodilian taxa we examined, the crocodile had the highest observed genetic diversity, with about three heterozygous sites per 10 kb. The lower heterozygosity of the other two crocodilians examined here is interesting given their recent or current status as endangered species. The gharial is critically endangered because of habitat loss (*41*), and the American alligator recently survived an anthropogenic population bottleneck (*42*) and was removed from the endangered species list in 1987. We inferred the effective population sizes of the alligator, crocodile, and gharial (Fig. 5A) using the neutral mutation rate for crocodilians ($\mu = 7.9 \times 10^{-9}$ substitutions site^{-1} generation^{-1}) calculated from the pairwise divergence (*17*) between the alligator and the saltwater crocodile. We note that the alligator and crocodile were

wild caught and thus were likely to represent the genetic diversity of their respective species, whereas the gharial we sequenced was bred in captivity and of unknown recent ancestry.

The crocodilians comprise many of the largest extant ectothermic species. As such, their success through recent geologic time is of special interest. Given their long generation time and slow mutation rate, the pairwise sequential Markovian coalescent (PSMC) model (*43*) approach can probe population sizes further into the past than is possible for faster-evolving lineages. Using this model, we found that all three lineages experienced distinct changes in their estimated effective population size (N_e) over the past 7 million years (Fig. 5B and fig. S26). We also included estimates of air temperature data (*44*) to identify any potential relationship of demographic histories to climate change. The results indicate that the crocodile and gharial both maintained relatively stable population sizes through the Pleistocene and Pliocene but both experienced sharp declines during the last cooling cycle, between ~100,000 and 10,000 years before the present (Fig. 5, B and C). In contrast, the population size of alligators declined continuously throughout the Pleistocene, perhaps because they inhabit more temperate latitudes and experienced greater effects from global cooling. A generally declining effective population size over the past million years was also shown for the Chinese alligator (*15*) using the PSMC approach.

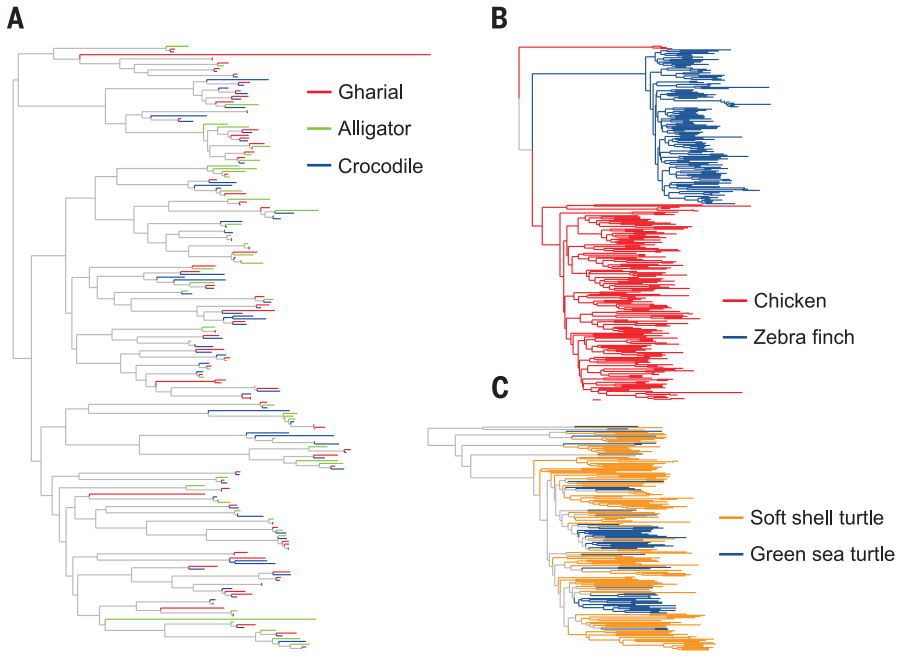


Fig. 4. OR expansions and contractions within archosaurs. Subtrees from neighbor-joining phylogenies of the intact crocodilian (**A**), avian (**B**), and testudine (**C**) OR repertoires. Crocodiles are represented by the gharial, American alligator, and saltwater crocodile; birds are represented by the chicken and zebra finch; and testudines are represented by softshell and green sea turtles. Note the paucity of lineage-specific (colored) clades among crocodilian ORs relative to avian and testudine ORs. Most crocodilian ORs are outparalogs (groups of paralogous genes that emerged prior to the divergence of the species analyzed), whereas the vast majority of avian and testudine ORs fall on monophyletic groups of inparalogs (groups of paralogous genes the emerged after the divergence of the species analyzed). Neighbor-joining trees were inferred using MEGA v5, a Poisson model of substitution; 1000 bootstrap iterations were performed to evaluate support. See also fig. S24.

A draft archosaur genome

One exciting use of genome sequence spanning archosaurs is the potential to infer the ancestral archosaur genome. As part of the WGA analysis, we computationally inferred the ancestral archosaur genome, along with ancestral genomes for all the internal nodes of the tree. Because of the constant turnover of sequence during the ~300 million years since the divergence of birds and crocodilians and the likelihood that some data are missing in the assemblies of extant taxa, the reconstructed genome assembly is limited to 584 Mb of sequence, less than the genome assemblies for extant taxa. Using a standard continuous time substitution model to determine the nucleotide at each position in the

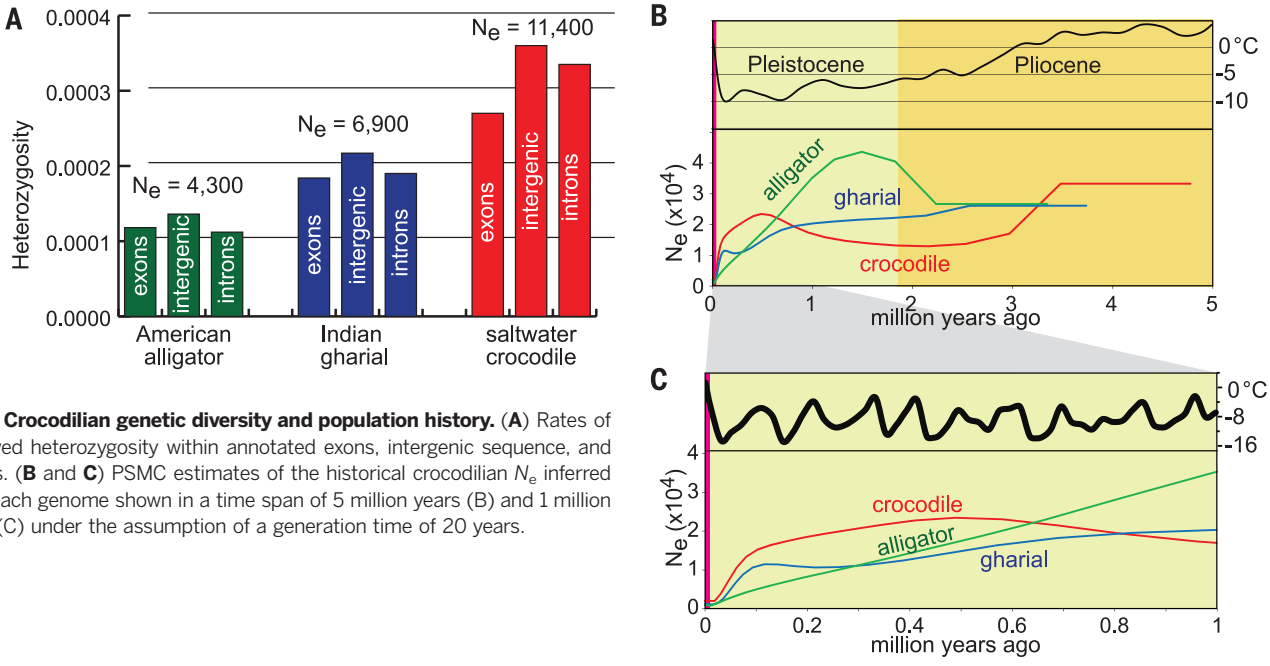


Fig. 5. Crocodilian genetic diversity and population history. (**A**) Rates of observed heterozygosity within annotated exons, intergenic sequence, and introns. (**B** and **C**) PSMC estimates of the historical crocodilian N_e inferred from each genome shown in a time span of 5 million years (B) and 1 million years (C) under the assumption of a generation time of 20 years.

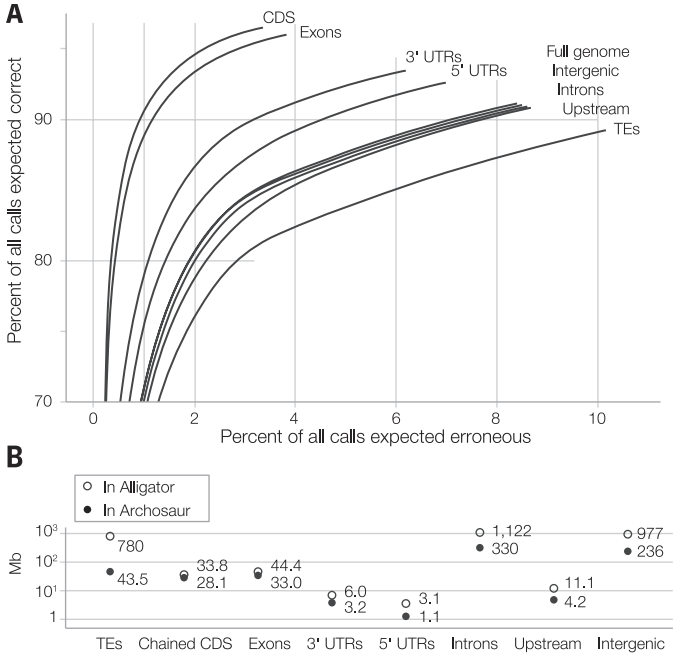


Fig. 6. Analyzing the archosaur assembly. (**A**) Expected base reconstruction accuracy. (**B**) Total archosaur bases assembled in several annotated functional classes and numbers of bases in each category from the alligator genome.

ancestral archosaur genome, the average expected reconstruction accuracy of archosaur bases is 91% (Fig. 6A and fig. S17).

The ancestral genome reconstruction exhibits a strong bias toward the recovery of functional elements. For example, we mapped alligator regions with various annotations, TEs, coding DNA sequences (CDSs), 3′ and 5′ untranslated regions (UTRs), exons, upstream sequences (defined by a 500-bp window upstream of the putative transcription start site for each gene), and introns to the archosaur genome, using the WGA to map the annotations by projection through the alignment. Relative to putatively neutrally evolving elements such as TEs, we found CDSs, 3′ UTRs, and 5′ UTRs (in decreasing order) to have substantially higher base-level reconstruction accuracy (e.g., 97% of base calls mapped by CDS annotations are expected to be correct; Fig. 6A). Concordantly, while on average only 26% of alligator bases had an aligned base in the archosaur reconstruction, the proportion of annotated bases mapping to archosaur was higher (Fig. 6B) (*17*). The reconstruction bias toward functional elements is correlated with differences in purifying selection as measured with PhyloP on the WGA (*17*). Transcribed elements annotated in alligator or chicken are also more likely to have remained stably ordered and oriented mapping back to the archosaur, which suggests that intragene ordering constraints have helped to preserve sequence structure (figs. S17 to S19).

Discussion

The draft genome assemblies of these three crocodilian taxa add to the growing list of available reptilian genomes and allow a more comprehensive analysis of vertebrate genome evolution. Because crocodilians are the sister group of birds, these three genomes also provide a critical resource for examining the ancestral state of various genomic features for birds, for which multiple genomes are now available (*45*). The most striking of our results is the remarkably low rate of genome-wide molecular evolution among all major crocodilian lineages. This low rate was observed for the accumulation of base substitutions at many different types of sites (those in UCE-anchored loci, 4D sites in protein-coding regions, and the presumably neutral sites in TE insertions) and for other types of genomic changes, such as micro-indels and TE movement. Recent genomic analyses of turtles suggest a low rate of evolution in that lineage as well (*13*), a finding we confirmed and extended. Taken as a whole, this provides strong evidence that a slow rate of genomic change is the ancestral state for archosaurs.

Our evidence that the low rate of molecular evolution applies to multiple types of genomic changes makes it tempting to speculate that there is a single underlying cause. Within mammals, the accelerated rate of molecular evolution for rodents relative to primates (also observed here; Fig. 1) is often attributed to shorter generation times along the rodent lineage (*46*). However, there have also been suggestions that the high rate in rodents could reflect differences in DNA repair efficiency (*47*). More broadly, rates of molecular evolution may be correlated with a number of factors, including body size and metabolic rate (*48*, *49*). However, these and other life history characters are themselves correlated (*50*, *51*), making it very difficult to untangle the relevant causal factors.

Our analyses include all major amniote lineages, and it is clear that crocodilians and turtles exhibit the lowest rates of molecular evolution; both of those clades are characterized by long generation times. Indeed, using a 20-year generation time along the crocodilian lineage (*17*), the inferred rate of molecular evolution per site per generation (7.9×10^{-9} substitutions per site per generation) is not substantially different from estimates in other lineages; it is the rate per year that is much lower for crocodilians. The higher rate for stem birds, which is actually similar to that observed for extant birds, could indicate that this lineage had already decreased their generation time. Indeed, recent analyses of paleontological data are highly consistent with decreased body size on the lineage ancestral to extant birds (*52*). Given the strong correlation between body size and generation time (*51*), this would be consistent with our observed changes in the average rate of molecular evolution. It will be of substantial interest to establish whether similar morphological correlates can be established for stem crocodilians and other lineages.

Materials and Methods

Sequencing and assembly

Genomic DNA was isolated using blood from four individuals: two *A. mississippiensis* and one each of *C. porosus* and *G. gangeticus*. Sequencing depth and assembly strategies differed depending on legacy data available for each taxon (*17*). Briefly, alligator data consisted of Illumina sequences from five libraries ranging from 5.5× to 88.7× coverage. These reads were assembled using AllPaths-LG (*53*) with default parameters. Legacy data from 21 fully sequenced BACs, 1309 BAC-end read pairs (*54*), and RNA-seq data, described below, were also used to aid the assembly. Crocodile data consisted of Illumina reads from three libraries ranging from 21.6× to 90.2× coverage. AllPaths-LG was used to assemble the raw data. As with the alligator genome draft, sequences from 360 major histocompatibility complex region BAC assemblies as well as RNA-seq data were used to aid the assembly. The gharial genome was assembled using SOAPdenovo v2.04 (*55*) and data from four Illumina libraries ranging from 50× to 170× coverage. No legacy data were available to improve the gharial assembly.

Transcriptome sequencing and sequence annotation

Total RNA was extracted from multiple alligator and crocodile tissues as well as gharial whole blood (*17*). RNA was extracted and subjected to library preparation and Illumina RNA-seq. While variable, most libraries had insert sizes between 300 and 350 bp and were sequenced both individually and as pools. In total, 11 Gb of high-quality sequence data were generated.

Gene predictions were made using Augustus (version 2.5.5) (*56*). RNA-seq data from *A. mississippiensis* were aligned to the draft American alligator genome with Tophat version 2.0.6 (*57*) and Bowtie version 2.0.5 (*58*). Augustus used these alignments to improve its gene predictions. Protein-coding genes predicted for alligator were then aligned to the other crocodilian assemblies with Genblastg version 1.38, and those alignments were used by Augustus to improve the gene predictions for those species. Functional annotation was accomplished by assigning gene nomenclature, GO, and pathway

information. Gene names were assigned on the basis of orthology or homology to species with a gene nomenclature project by transferring names to the crocodilian genes. GO was assigned to predicted proteins according to a combinatorial approach (*17*). Pathway information was assigned on the basis of reciprocal BLAST. Annotated genes, gene products, and genome assemblies are available at NCBI, CrocBase (http://crocgenome.hpc.msstate.edu/crocbase/gene.php), and via the Comparative Genomics (CoGe) browser (http://genomevolution.org/CoGe).

TEs in the genomes were identified and annotated by three laboratories semi-independently. Briefly, TEs were identified de novo in a given genome draft with either RepeatModeler (*59*) or a combination of PILER (*60*), RepeatScout (*61*), and LTRHarvest (*62*). Output from each method was curated using a combination of manual inspection and computational tools. Combining TE consensus sequences from all three crocodilians resulted in a library of 1269 different TEs. Full details of all sequence annotation protocols are in (*17*).

UCE identification and analysis

To create a large set of UCE loci (*17*), we combined two sets of UCEs (*25*, *63*) and kept unique and nonduplicate loci in the set (n = 8047 UCE loci). Using the positions of these loci in the chicken genome (galGal3), we designed capture probes (n = 12,237) for each locus to use for in silico identification of orthologous UCEs in other tetrapods (table S6) and aligned each capture probe to those genomes. After identification of putative UCE loci in each genome, we sliced the match location of all probes ± 2000 bp from each genome assembly and recovered slices derived from multiple probes targeting the same locus, then reassembled sequences back into full UCE loci. We then trimmed all slices to approximately the length of the UCE locus ± 1000 bp and identified the set of all loci found in all taxa (a complete matrix) from two different taxon samples (table S8). We named these taxon-set-1 and taxon-set-2. Taxon-set-1 includes the Western clawed frog (*Xenopus tropicalis*) and consequently contains fewer orthologous loci in a complete matrix.

Using the complete data matrices, we aligned FASTA data corresponding to each reassembled UCE locus for each taxon. After alignment and trimming, we removed any loci containing ambiguous base calls. The remaining alignment data for taxon-set-1 contained 604 loci totaling 495,744 characters and 93,374 alignment patterns [mean locus length = 820 bp; 95% confidence interval (CI) = 47 bp]. The remaining alignment data for taxon-set-2 contained 965 loci totaling 878,786 characters and 172,112 alignment patterns (mean locus length = 911 bp; 95% CI = 40 bp). We concatenated all loci in each set, and we analyzed the resulting concatenated alignments with RAxML 7.3.4 (*64*), conducting 20 maximum likelihood (ML) tree searches and 500 bootstrap replicates for each data set. Using RAxML, we checked for bootstrap replicate convergence using the "autoMRE" function. Both data sets converged after 50 replicates, and we used RAxML to reconcile each best ML tree with each set of 500 bootstrap replicates. We also conducted partitioned, concatenated analyses of the UCE data, but these results did not differ from the unpartitioned results (*17*).

Phylome analysis

Complete collections of ML gene trees for every gene encoded in each of the three crocodilian genomes (phylomes) were reconstructed using the phylomeDB pipeline (*17*, *65*). In brief, sequence searches were used to retrieve homologs (E-value 10^{-5}, 50% overlap) in a set of vertebrates (*17*). These were aligned using three different programs in forward and reverse orientation. Consensus alignments were built with T-coffee (*66*) and trimmed with trimAl (*67*). The evolutionary model best fitting the data was used to build an ML tree with PhyML (*68*) using four rate categories and a fraction of invariable sites, estimated from the data. Branch support was computed using an aLRT (approximate likelihood ratio test) parametric test. Orthology and paralogy relationships among crocodilian genes and those encoded by the other genomes were inferred from the phylomes, using a species-overlap algorithm (*69*) as implemented in ETE (*70*). The resulting trees and orthology and paralogy predictions can be accessed through phylomeDB.org (*19*). The crocodilian phylomes were scanned to detect and date duplication events using a previously described algorithm (*71*). For species tree reconstruction, two complementary approaches were used. First, a supertree was inferred from all trees in the three phylomes by means of a gene tree parsimony approach, as implemented in the dup-tree algorithm (*72*). Second, the alignments of 337 gene families with one-to-one orthology in all considered species were concatenated and used to build a ML phylogeny as described above.

Gene family analysis

We conducted bioinformatic searches to characterize the repertoires of ORs, vomeronasal receptors (V1R and V2R), taste receptors (T1R and T2R), and trace amine-associated receptors (TAAR) of the three crocodilians in our study, and we compared the repertoires with representative vertebrates (table S21) (*17*). We focused the majority of our analyses on the ORs. Briefly, we performed TblastN searches of the three crocodilian genomes using known vertebrate ORs as queries, and the best nonoverlapping BLAST hits were extracted. Putative complete OR genes were added to the amino acid query, and a new TBlastN search was conducted to annotate pseudogenes and truncated genes. Putative ORs were annotated to their subfamily by comparing amino acid sequences against a BLASTP database of known OR amino acid sequences. Phylogenetic analyses were conducted using MEGA v5 (*73*). We inferred neighbor-joining phylogenies to assess patterns of divergence and diversity of intact crocodilian ORs relative to other vertebrates using a Poisson model of substitution and evaluated support for the nodes with 1000 pseudoreplicates. We compared the evolution of ORs for the three crocodilians, chicken, and zebra finch (*74*) as well as green sea turtle and Asian softshell turtle (*16*).

Genome alignments and ancestral genome reconstruction

The WGA of 23 taxa (table S12) (*17*) was computed using progressive-cactus (github.com/glennhickey/progressiveCactus) with default parameters and the phylogeny shown in Fig. 2A (*75*). The topology of the phylogeny was derived by manually merging a subtree of the UCE tree (*17*) with results from the accompanying avian phylogeny paper (*76*) along with published phylogenies for passerine birds (*77*), parrots (*78*), and turtles (*79*). Nucleotide-level ancestral reconstruction of all internal nodes was performed as part of the process, using a phylogenetically weighted form of the algorithm described in Nguyen *et al.* (*80*) and appropriate for partial genome assemblies. To improve the ancestral base calls, we used the ancestorsML tool in the HAL tools library (github.com/glennhickey/hal) (*81*) to call bases by ML, using the general reversible continuous-time nucleotide substitution model. To parameterize the model and estimate branch lengths, we used phyloFit (*82*) on conserved 4D sites in alligator genes (*17*). A complete technical exposition of the alignment computation and statistics calculated is available in (*17*).

Mutation rate estimation

We used a phylogenetic approach to estimate the overall mutation rate μ along the crocodilian lineage. From both the WGA between alligator and crocodile and the multiple sequence alignment that includes alligator and crocodile, we estimate the overall divergence between alligator and crocodile to be 7.1%. Because of the remarkably small divergence between these two, we assumed an infinite-sites model of evolution and ignored back-mutations. To calculate a per-generation mutation rate, we used 90 Ma as the TMRCA of alligator and crocodile and an average generation time of 20 years (table S23) (*17*).

Heterozygosity and population history estimation

For each genome, we used BWA (*83*) to map paired-end genome reads from a single individual back to the final genome assembly (*17*). We used tools in the GATK package (www.broadinstitute.org/gatk) to perform indel realignment of each read around possible insertion-deletion positions, then analyzed all genomic positions where the read depth was exactly equal to the genome-wide mean. We derived cutoffs to distinguish bona fide heterozygous positions from sequence error by analysis of mutation spectra at these sites (table S24). From this analysis, we calculated the observed rate of heterozygosity H at intergenic sequence in each species: alligator H = 0.000136, gharial H = 0.000217, and crocodile H = 0.000360. Using these values as an estimate for θ and the substitution rate m calculated above, we estimated the effective population size for each species as shown in Fig. 6.

To estimate historical population sizes, we called single-nucleotide polymorphisms with SAMtools using reads with a map score of >30 and base calls with a quality score of >20. We applied the PSMC (*43*) model using 20 years for the generation time (table S23). We used 90 Ma as the TMRCA of *C. porosus* and *A. mississippiensis*, and our analyses indicate 7.1% divergence. Therefore, given a 20-year generation time, we calculated a mutation rate of 7.89×10^{-9} year^{-1} site^{-1}. We conducted bootstrap tests for each of the three taxa by splitting the scaffolds into smaller segments and randomly sampling the segments with replacement (fig. S26). We used 100 replicates to test the robustness of the returned population demographic history. We also gathered ancestral Northern Hemisphere air temperature data from (*44*) and took averages for 200,000-year bins. Climate oscillations over the past 1 million years were calculated in 20,000-year bins.

REFERENCES AND NOTES

1. A. Janke, U. Arnason, The complete mitochondrial genome of *Alligator mississippiensis* and the separation between recent archosauria (birds and crocodiles). *Mol. Biol. Evol.* **14**, 1266–1272 (1997). doi: 10.1093/oxfordjournals.molbev.a025736; pmid: 9402737
2. A. G. Sennikov, The first ctenosauriscid (Reptilia: Archosauromorpha) from the Lower Triassic of eastern Europe. *Paleontol. J.* **46**, 499–511 (2012). doi: 10.1134/S0031030112050097
3. C. A. Brochu, Phylogenetic approaches toward crocodylian history. *Annu. Rev. Earth Planet. Sci.* **31**, 357–397 (2003). doi: 10.1146/annurev.earth.31.100901.141308
4. G. Grigg, F. Seebacher, C. E. Franklin, Eds., *Crocodilian Biology and Evolution* (Surrey Beatty, Chipping Norton, NSW, Australia, 2001).
5. M. M. Cohen, C. Gans, The chromosomes of the order Crocodilia. *Cytogenetics* **9**, 81–105 (1970). doi: 10.1159/000130080; pmid: 5461083
6. J. W. Lang, H. V. Andrews, Temperature-dependent sex determination in crocodilians. *J. Exp. Zool.* **270**, 28–44 (1994). doi: 10.1002/jez.1402700105
7. E. Olmo, Evolution of genome size and DNA base composition in reptiles. *Genetica* **57**, 39–50 (1981). doi: 10.1007/BF00057541
8. J. Harshman, C. J. Huddleston, J. P. Bollback, T. J. Parsons, M. J. Braun, True and false gharials: A nuclear gene phylogeny of crocodylia. *Syst. Biol.* **52**, 386–402 (2003). doi: 10.1080/10635150390197028; pmid: 12775527
9. J. R. Oaks, A time-calibrated species tree of Crocodylia reveals a recent radiation of the true crocodiles. *Evolution* **65**, 3285–3297 (2011). doi: 10.1111/j.1558-5646.2011.01373.x; pmid: 22023592
10. A. Janke, A. Gullberg, S. Hughes, R. K. Aggarwal, U. Arnason, Mitogenomic analyses place the gharial (*Gavialis gangeticus*) on the crocodile tree and provide pre-K/T divergence times for most crocodilians. *J. Mol. Evol.* **61**, 620–626 (2005). doi: 10.1007/s00239-004-0336-9; pmid: 16211427
11. J. Alföldi *et al*., The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature* **477**, 587–591 (2011). doi: 10.1038/nature10390; pmid: 21881562
12. T. A. Castoe *et al*., The Burmese python genome reveals the molecular basis for extreme adaptation in snakes. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 20645–20650 (2013). pmid: 24297902
13. H. B. Shaffer *et al*., The western painted turtle genome, a model for the evolution of extreme physiological adaptations in a slowly evolving lineage. *Genome Biol.* **14**, R28 (2013). doi: 10.1186/gb-2013-14-3-r28; pmid: 23537068
14. F. J. Vonk *et al*., The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 20651–20656 (2013). doi: 10.1073/pnas.1314702110; pmid: 24297900
15. Q. H. Wan *et al*., Genome analysis and signature discovery for diving and sensory properties of the endangered Chinese alligator. *Cell Res.* **23**, 1091–1105 (2013). doi: 10.1038/cr.2013.104; pmid: 23917531
16. Z. Wang *et al*., The draft genomes of soft-shell turtle and green sea turtle yield insights into the development and evolution of the turtle-specific body plan. *Nat. Genet.* **45**, 701–706 (2013). doi: 10.1038/ng.2615; pmid: 23624526
17. See supplementary materials on *Science* Online.
18. L. W. Hillier *et al*., Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695–716 (2004). doi: 10.1038/nature03154; pmid: 15592404
19. J. Huerta-Cepas, S. Capella-Gutiérrez, L. P. Pryszcz, M. Marcet-Houben, T. Gabaldón, PhylomeDB v4: Zooming into the plurality of evolutionary histories of a genome. *Nucleic Acids Res.* **42**, D897–D902 (2014). doi: 10.1093/nar/gkt1177; pmid: 24275491
20. E. S. Lander *et al*., Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001). doi: 10.1038/35057062; pmid: 11237011
21. W. C. Warren *et al*., The genome of a songbird. *Nature* **464**, 757–762 (2010). doi: 10.1038/nature08819; pmid: 20360741
22. W. C. Warren *et al*., Genome analysis of the platypus reveals unique signatures of evolution. *Nature* **453**, 175–183 (2008). doi: 10.1038/nature06936; pmid: 18464734
23. A. T. Chinwalla *et al*., Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002). doi: 10.1038/nature01262; pmid: 12466850
24. G. Bejerano *et al*., Ultraconserved elements in the human genome. *Science* **304**, 1321–1325 (2004). doi: 10.1126/science.1098119; pmid: 15131266
25. B. C. Faircloth *et al*., Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Syst. Biol.* **61**, 717–726 (2012). doi: 10.1093/sysbio/sys004; pmid: 22232343
26. J. E. McCormack *et al*., Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species-tree analysis. *Genome Res.* **22**, 746–754 (2012). doi: 10.1101/gr.125864.111; pmid: 22207614
27. A. Siepel *et al*., Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005). doi: 10.1101/gr.3715005; pmid: 16024819
28. Y. Chiari, V. Cahais, N. Galtier, F. Delsuc, Phylogenomic analyses support the position of turtles as the sister group of birds and crocodiles (Archosauria). *BMC Biol.* **10**, 65 (2012). doi: 10.1186/1741-7007-10-65; pmid: 22839781
29. N. G. Crawford *et al*., More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs. *Biol. Lett.* **8**, 783–786 (2012). doi: 10.1098/rsbl.2012.0331; pmid: 22593086
30. A. C. Tzika, R. Helaers, G. Schramm, M. C. Milinkovitch, Reptilian-transcriptome v1.0, a glimpse in the brain transcriptome of five divergent Sauropsida lineages and the phylogenetic position of turtles. *EvoDevo* **2**, 19 (2011). doi: 10.1186/2041-9139-2-19; pmid: 21943375
31. J. V. Chamary, J. L. Parmley, L. D. Hurst, Hearing silence: Non-neutral evolution at synonymous sites in mammals. *Nat. Rev. Genet.* **7**, 98–108 (2006). doi: 10.1038/nrg1770; pmid: 16418745
32. A. Künstner, B. Nabholz, H. Ellegren, Significant selective constraint at 4-fold degenerate sites in the avian genome and its consequence for detection of positive selection. *Genome Biol. Evol.* **3**, 1381–1389 (2011). doi: 10.1093/gbe/evr112; pmid: 22042333
33. C. A. Brochu, Morphology, fossils, divergence timing, and the phylogenetic relationships of Gavialis. *Syst. Biol.* **46**, 479–522 (1997). doi: 10.1093/sysbio/46.3.479; pmid: 11975331
34. G. V. Glazko, M. Nei, Estimation of divergence times for major lineages of primate species. *Mol. Biol. Evol.* **20**, 424–434 (2003). doi: 10.1093/molbev/msg050; pmid: 12644563
35. R. Cordaux, J. Lee, L. Dinoso, M. A. Batzer, Recently integrated Alu retrotransposons are essentially neutral residents of the human genome. *Gene* **373**, 138–144 (2006). doi: 10.1016/j.gene.2006.01.020; pmid: 16527433
36. B. Paten *et al*., Genome-wide nucleotide-level mammalian ancestor reconstruction. *Genome Res.* **18**, 1829–1843 (2008). doi: 10.1101/gr.076521.108; pmid: 18849525
37. M. Blanchette, E. D. Green, W. Miller, D. Haussler, Reconstructing large regions of an ancestral mammalian genome in silico. *Genome Res.* **14**, 2412–2423 (2004). doi: 10.1101/gr.2800104; pmid: 15574820
38. Y. Niimura, Olfactory receptor multigene family in vertebrates: From the viewpoint of evolutionary genomics. *Curr. Genomics* **13**, 103–114 (2012). doi: 10.2174/138920212799860706; pmid: 23024602
39. J. Huerta-Cepas, H. Dopazo, J. Dopazo, T. Gabaldón, The human phylome. *Genome Biol.* **8**, R109 (2007). pmid: 17567924
40. Y. Niimura, M. Nei, Extensive gains and losses of olfactory receptor genes in mammalian evolution. *PLOS ONE* **2**, e708 (2007). pmid: 17684554
41. C. Stevenson, R. Whitaker, in *Crocodiles. Status Survey and Conservation Action Plan*, S. C. Manolis, C. Stevenson, Eds. (Crocodile Specialist Group, Darwin, Australia, 2010), pp. 139–143.
42. R. M. Elsey, A. R. Woodward, in *Crocodiles. Status Survey and Conservation Action Plan*, S. C. Manolis, C. Stevenson, Eds. (Crocodile Specialist Group, Darwin, Australia, 2010), pp. 1–4.
43. H. Li, R. Durbin, Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493–496 (2011). doi: 10.1038/nature10231; pmid: 21753753
44. R. S. W. van de Wal, B. de Boer, L. J. Lourens, P. Kohler, R. Bintanja, Reconstruction of a continuous high-resolution CO_2 record over the past 20 million years. *Clim. Past* **7**, 1459–1469 (2011). doi: 10.5194/cp-7-1459-2011
45. G. Zhang *et al*., Comparative genomics reveals insights into avian genome evolution and adaptation. *Science* **346**, 1311–1320 (2014).
46. C. I. Wu, W. H. Li, Evidence for higher rates of nucleotide substitution in rodents than in man. *Proc. Natl. Acad. Sci. U.S.A.* **82**, 1741–1745 (1985). doi: 10.1073/pnas.82.6.1741; pmid: 3856856
47. Y. H. Lin, P. J. Waddell, D. Penny, Pika and vole mitochondrial genomes increase support for both rodent monophyly and glires. *Gene* **294**, 119–129 (2002). doi: 10.1016/S0378-1119(02)00695-9; pmid: 12234673
48. A. P. Martin, S. R. Palumbi, Body size, metabolic rate, generation time, and the molecular clock. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 4087–4091 (1993). doi: 10.1073/pnas.90.9.4087; pmid: 8483925
49. J. F. Gillooly, A. P. Allen, G. B. West, J. H. Brown, The rate of DNA evolution: Effects of body size and temperature on the molecular clock. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 140–145 (2005). doi: 10.1073/pnas.0407735101; pmid: 15618408
50. L. Blueweiss *et al*., Relationships between body size and some life history parameters. *Oecologia* **37**, 257–272 (1978). doi: 10.1007/BF00344996
51. V. M. Savage, J. F. Gillooly, J. H. Brown, G. B. West, E. L. Charnov, Effects of body size and temperature on population growth. *Am. Nat.* **163**, 429–441 (2004). doi: 10.1086/381872; pmid: 15026978
52. M. S. Lee, A. Cau, D. Naish, G. J. Dyke, Sustained miniaturization and anatomical innovation in the dinosaurian ancestors of birds. *Science* **345**, 562–566 (2014). doi: 10.1126/science.1252243; pmid: 25082702
53. S. Gnerre *et al*., High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 1513–1518 (2011). doi: 10.1073/pnas.1017351108; pmid: 21187386
54. A. M. Shedlock *et al*., Phylogenomics of nonavian reptiles and the structure of the ancestral amniote genome. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 2767–2772 (2007). doi: 10.1073/pnas.0606204104; pmid: 17307883
55. R. Li *et al*., De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res.* **20**, 265–272 (2010). doi: 10.1101/gr.097261.109; pmid: 20019144
56. M. Stanke, M. Diekhans, R. Baertsch, D. Haussler, Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**, 637–644 (2008). doi: 10.1093/bioinformatics/btn013; pmid: 18218656
57. D. Kim *et al*., TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013). doi: 10.1186/gb-2013-14-4-r36; pmid: 23618408
58. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009). doi: 10.1186/gb-2009-10-3-r25; pmid: 19261174
59. A. F. A. Smit, R. Hubley, RepeatModeler Open-1.0; www.repeatmasker.org.
60. R. C. Edgar, E. W. Myers, PILER: Identification and classification of genomic repeats. *Bioinformatics* **21** (suppl. 1), i152–i158 (2005). doi: 10.1093/bioinformatics/bti1003; pmid: 15961452
61. A. L. Price, N. C. Jones, P. A. Pevzner, De novo identification of repeat families in large genomes. *Bioinformatics* **21** (suppl. 1), i351–i358 (2005). doi: 10.1093/bioinformatics/bti1018; pmid: 15961478

62. D. Ellinghaus, S. Kurtz, U. Willhoeft, LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC Bioinformatics* **9**, 18 (2008). doi: 10.1186/1471-2105-9-18; pmid: 18194517
63. S. Stephen, M. Pheasant, I. V. Makunin, J. S. Mattick, Large-scale appearance of ultraconserved elements in tetrapod genomes and slowdown of the molecular clock. *Mol. Biol. Evol.* **25**, 402–408 (2008). doi: 10.1093/molbev/msm268; pmid: 18056681
64. A. Stamatakis, RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006). doi: 10.1093/bioinformatics/btl446; pmid: 16928733
65. J. Huerta-Cepas, A. Bueno, J. Dopazo, T. Gabaldón, PhylomeDB: A database for genome-wide collections of gene phylogenies. *Nucleic Acids Res.* **36**, D491–D496 (2008). doi: 10.1093/nar/gkm899; pmid: 17962297
66. M. R. Wallace *et al*., A de novo Alu insertion results in neurofibromatosis type 1. *Nature* **353**, 864–866 (1991). doi: 10.1038/353864a0; pmid: 1719426
67. S. Capella-Gutiérrez, J. M. Silla-Martínez, T. Gabaldón, trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972–1973 (2009). doi: 10.1093/bioinformatics/btp348; pmid: 19505945
68. S. Guindon *et al*., New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. *Syst. Biol.* **59**, 307–321 (2010). doi: 10.1093/sysbio/syq010; pmid: 20525638
69. T. Gabaldón, Large-scale assignment of orthology: Back to phylogenetics? *Genome Biol.* **9**, 235 (2008). doi: 10.1186/gb-2008-9-10-235; pmid: 18983710
70. J. Huerta-Cepas, J. Dopazo, T. Gabaldón, ETE: A python environment for tree exploration. *BMC Bioinformatics* **11**, 24 (2010). doi: 10.1186/1471-2105-11-24; pmid: 20070885
71. J. Huerta-Cepas, T. Gabaldón, Assigning duplication events to relative temporal scales in genome-wide studies. *Bioinformatics* **27**, 38–45 (2011). doi: 10.1093/bioinformatics/btq609; pmid: 21075746
72. A. Wehe, M. S. Bansal, J. G. Burleigh, O. Eulenstein, DupTree: A program for large-scale phylogenetic analyses using gene tree parsimony. *Bioinformatics* **24**, 1540–1541 (2008). doi: 10.1093/bioinformatics/btn230; pmid: 18474508
73. K. Tamura *et al*., MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739 (2011). doi: 10.1093/molbev/msr121; pmid: 21546353
74. S. S. Steiger, V. Y. Kuryshev, M. C. Stensmyr, B. Kempenaers, J. C. Mueller, A comparison of reptilian and avian olfactory receptor gene repertoires: Species-specific expansion of group gamma genes in birds. *BMC Genomics* **10**, 446 (2009). doi: 10.1186/1471-2164-10-446; pmid: 19772566
75. B. Paten *et al*., Cactus: Algorithms for genome multiple sequence alignment. *Genome Res.* **21**, 1512–1528 (2011). doi: 10.1101/gr.123356.111; pmid: 21665927
76. E. D. Jarvis *et al*., Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* **346**, 1320–1331 (2014).
77. F. K. Barker, A. Cibois, P. Schikler, J. Feinstein, J. Cracraft, Phylogeny and diversification of the largest avian radiation. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 11040–11045 (2004). doi: 10.1073/pnas.0401892101; pmid: 15263073
78. T. F. Wright *et al*., A multilocus molecular phylogeny of the parrots (Psittaciformes): Support for a Gondwanan origin during the Cretaceous. *Mol. Biol. Evol.* **25**, 2141 (2008). doi: 10.1093/molbev/msn160
79. R. C. Thomson, H. B. Shaffer, Sparse supermatrices for phylogenetic inference: Taxonomy, alignment, rogue taxa, and the phylogeny of living turtles. *Syst. Biol.* **59**, 42–58 (2010). doi: 10.1093/sysbio/syp075; pmid: 20525619
80. N. Nguyen *et al*., Building a pangenome reference for a population. *Res. Comput. Mol. Biol.* **8394**, 207–221 (2014). doi: 10.1007/978-3-319-05269-4_17
81. G. Hickey, B. Paten, D. Earl, D. Zerbino, D. Haussler, HAL: A hierarchical format for storing and analyzing multiple genome alignments. *Bioinformatics* **29**, 1341–1342 (2013). doi: 10.1093/bioinformatics/btt128; pmid: 23505295
82. M. J. Hubisz, K. S. Pollard, A. Siepel, PHAST and RPHAST: Phylogenetic analysis with space/time models. *Brief. Bioinform.* **12**, 41–51 (2011). doi: 10.1093/bib/bbq072; pmid: 21278375
83. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009). doi: 10.1093/bioinformatics/btp324; pmid: 19451168

ACKNOWLEDGMENTS

Genome drafts have been submitted to the National Center for Biotechnology Information repository under the following accession numbers: *Alligator mississippiensis*, AKHW00000000; *Crocodylus porosus*, JRXG00000000; *Gavialis gangeticus*, JRWT00000000. Supplemental files have been archived at GigaScience (DOIs: 10.5524/100125, 10.5524/100126, 10.5524/100127, and 10.5524/100128) and at crocgenomes.org. This project was conducted by the International Crocodilian Genomes Working Group (ICGWG; www.crocgenomes.org). Supported by NSF grants MCB-1052500 and DEB-1020865 (D.A.R.), MCB-0841821 (D.A.R., D.G.P., F.M.M., C.J.S.), DUE-0920151 (E.L.B., E.W.T.), DBI-0905714 (M.K.F.), and DEB-1242260 (B.C.F.). D.A.R., F.M.M., and D.G.P. were also supported by the Institute for Genomics, Biocomputing and Biotechnology at Mississippi State University. S.R.I., L.G.M., J.G., and C.M. were supported by Australian Rural Industries Research and Development Corporation grants (RIRDC PRJ-000549, RIRDC PRJ- 005355, RIRDC PRJ-002461). R.E.G. is a Searle Scholar, Sloan Fellow, and consultant for Dovetail Genomics.. E.D.J. was supported by the Howard Hughes Medical Institute and the National Institutes of Health. E.L. received support from the Gordon and Betty Moore Foundation (#3383). R. Elsey, S. Lance, and T. Tuberville aided in collecting alligator samples. The following centers were vital in permitting the computational analyses required for this project: The High Performance Computing Collaborative (HPC2) at Mississippi State University, the High Performance Computing Center at Texas Tech University, the Georgia Advanced Computing Resource Center at the University of Georgia, the University of Florida High Performance Computing Center. The National Institutes of Health provided funding for UCSC infrastructure used in computing whole genome alignments and ancestral genome reconstructions (1U41HG007234-01, 1U41HG006992-2, 5U01HG004695). Finally, we are grateful to K. Vliet and D. Barber for providing access to fresh gharial blood.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/346/6215/1254449/suppl/DC1
Materials and Methods
Figs. S1 to S29
Tables S1 to S24
References (*84*–*206*)

7 April 2014; accepted 6 November 2014
10.1126/science.1254449

RESEARCH ARTICLE SUMMARY

AVIAN GENOMICS

Evidence for a single loss of mineralized teeth in the common avian ancestor

Robert W. Meredith,* Guojie Zhang, M. Thomas P. Gilbert, Erich D. Jarvis, Mark S. Springer*

INTRODUCTION: The absence of teeth or edentulism has evolved on multiple occasions within vertebrates, including birds, turtles, and a few groups of mammals (anteaters, baleen whales, and pangolins). There are also mammals with enamelless teeth (aardvarks, sloths, and armadillos). All toothless/enamelless vertebrates are descended from ancestors with enamel-capped teeth. In the case of birds, it is theropod dinosaurs. Instead of teeth, modern birds use a horny beak (rhamphotheca) and part of their digestive tract (muscular gizzard) to grind up and process food. The fossil record of early birds is fragmentary, and it is unclear whether tooth loss evolved in the common ancestor of all modern birds or convergently in two or more independent lineages.

RATIONALE: Tooth formation in vertebrates is a complicated process that involves many different genes. Of these genes, six are essential for the proper formation of dentin (*DSPP*) and enamel (*AMTN*, *AMBN*, *ENAM*, *AMELX*, and *MMP20*). We examined these six genes in the genomes of 48 bird species, which represent nearly all living bird orders, as well as the American alligator, a representative of Crocodylia (the closest living relatives of birds), for the presence of inactivating mutations that are shared by all 48 birds. The presence of such shared mutations in dentin and enamel-related genes would suggest a single loss of mineralized teeth in the common ancestor of all living birds. We also queried the genomes of additional toothless/enamelless vertebrates, including three turtles and four mammals, for inactivating mutations in these genes. For comparison, we looked at the genomes of mammalian taxa with enamel-capped teeth.

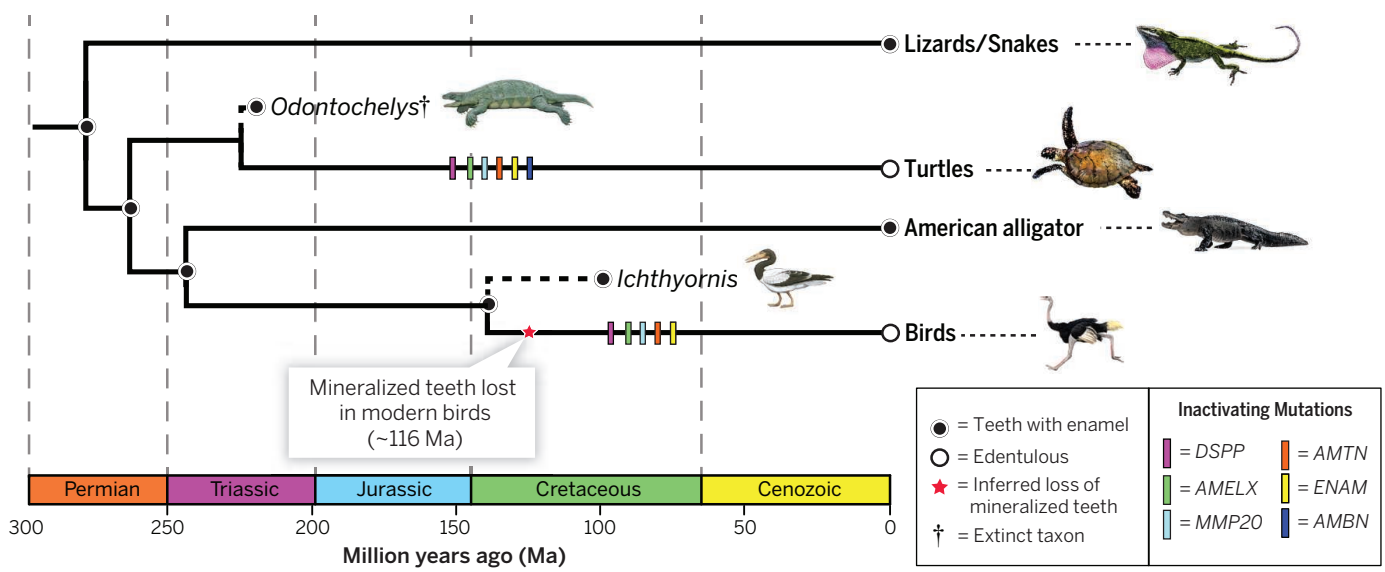
RESULTS: All edentulous vertebrate genomes that were examined are characterized by inactivating mutations in *DSPP*, *AMBN*, *AMELX*, *AMTN*, *ENAM*, and *MMP20*, rendering these genes nonfunctional. The dentin-related gene *DSPP* is functional in vertebrates with enamelless teeth (sloth, aardvark, and armadillo). All six genes are functional in the American alligator and mammalian taxa with enamel-capped teeth. More important, 48 bird species share inactivating mutations in both dentin-related (*DSPP*) and enamel-related genes (*ENAM*, *AMELX*, *AMTN*, and *MMP20*), indicating that the genetic machinery necessary for tooth formation was lost in the common ancestor of all modern birds. Furthermore, the frameshift mutation rate in birds suggests that the outer enamel covering of teeth was lost about 116 million years ago.

ON OUR WEB SITE

Read the full article at http://dx.doi.org/10.1126/science.1254390

CONCLUSIONS: We postulate, on the basis of fossil and molecular evidence, a two-step scenario whereby tooth loss and beak development evolved together in the common ancestor of all modern birds. In the first stage, tooth loss and partial beak development commenced on the anterior portion of both the upper and lower jaws. The second stage involved concurrent progression of tooth loss and beak development from the anterior portion of both jaws to the back of the rostrum. We propose that this progression ultimately resulted in a complete horny beak that effectively replaced the teeth and may have contributed to the diversification of living birds. ■

The list of author affiliations is available in the full article online.
*Corresponding author. E-mail: meredithr@mail.montclair.edu (R.W.M.); mark.springer@ucr.edu (M.S.S.)
Cite this article as R. W. Meredith *et al.*, *Science* **346**, 1254390 (2014). DOI: 10.1126/science.1254390



Observed shared inactivating mutations in tooth formation. Related genes were mapped onto a time tree depicting evolutionary relationships and times of divergence between modern birds, the closely related extinct taxon *Ichthyornis*, and the American alligator. The hypothesized loss of mineralized teeth on the modern bird branch at 116 million years ago (Ma) is based on frameshift mutation rates.

Published by AAAS

Downloaded from www.sciencemag.org on December 14, 2014

RESEARCH ARTICLE

Evidence for a single loss of mineralized teeth in the common avian ancestor

Robert W. Meredith,[1]* Guojie Zhang,[2,3] M. Thomas P. Gilbert,[4,5] Erich D. Jarvis,[6] Mark S. Springer[7]*

Edentulism, the absence of teeth, has evolved convergently among vertebrates, including birds, turtles, and several lineages of mammals. Instead of teeth, modern birds (Neornithes) use a horny beak (rhamphotheca) and a muscular gizzard to acquire and process food. We performed comparative genomic analyses representing lineages of nearly all extant bird orders and recovered shared, inactivating mutations within genes expressed in both the enamel and dentin of teeth of other vertebrate species, indicating that the common ancestor of modern birds lacked mineralized teeth. We estimate that tooth loss, or at least the loss of enamel caps that provide the outer layer of mineralized teeth, occurred about 116 million years ago.

With ~10,000 species, birds are the most species-rich clade of edentulous (toothless) jawed vertebrates (*1*). All edentulous or enamelless amniote lineages are derived from ancestors that possessed enamel-capped teeth (Fig. 1). Edentulism also occurs in turtles and mammalian lineages (anteaters, baleen whales, and pangolins) (*2*, *3*), with some mammals lacking tooth enamel (aardvarks and sloths) (*4*), but these clades are less diverse than birds and together comprise only ~385 species. Modern birds share a common ancestry with toothed, maniraptoran theropod dinosaurs (*5*). Instead of teeth, modern birds use two highly specialized structures, a rhamphotheca (horny beak) for food acquisition and, as part of the digestive tract, a muscular gizzard for food processing (*5*). These structures are proposed to have played a role in enabling birds to adapt to a multitude of previously unavailable ecospaces and diversify relative to other edentulous vertebrates (*5*).

Throughout their evolutionary history [~170 million years ago (Ma)], independent tooth reduction and loss has occurred in multiple lineages of Avialae (living birds and all fossil theropods that are more closely related to Neornithes than to the dinosaurs *Dromaeosaurus albertensis* or *Troodon formosus*) (*6*). Minimally, there are six lineages of Avialae that show evidence of tooth reduction, with four lineages exhibiting complete tooth loss. Neornithes are hypothesized to have become edentulous somewhere between 125 and 65.5 Ma (*5*, *7*). Among avialaen lineages exhibiting tooth reduction or loss, a rhamphotheca has also independently evolved (*5*, *7*). The multiple independent iterations of tooth reduction and loss in Avialae suggest that ancestral odontogenetic pathways may have been predisposed to tooth reduction and loss in this clade (*5*).

Most phylogenetic analyses suggest that teeth were lost in the common ancestor of modern birds (*8*, *9*), although dentate taxa from the Mesozoic (*Hesperornis* and *Ichthyornis*) have been recovered inside of crown Neornithes in some cladistic studies, implying that tooth loss could have instead occurred convergently among different members of this clade (*10*). The oldest fossils of putative neornithines are from the Cretaceous, but these fossils are fragmentary, lack complete jaw elements, and fail to resolve whether edentulism evolved in the common ancestor of all modern birds or convergently in at least two subclades (*5*).

Parallel tooth and enamel loss

The inactivation of enamel-related genes parallels the loss of enamel in the mammalian fossil record (*4*, *11*–*13*). The presence of shared, inactivating mutations in dentin- and enamel-related genes could therefore be used as a molecular proxy to assess whether teeth were lost once or on multiple occasions in the common ancestor of living birds, as well as to estimate the timing of tooth loss. We queried the genomes of 48 living bird species that span the phylogeny of modern Aves (*14*), including representatives of the two major clades of birds (Palaeognathae and Neognathae) and nearly all extant bird orders, as well as a representative of Crocodylia [*Alligator mississippiensis* (American alligator)], which is the extant sister group of birds, by performing a comparative analysis of six genes that have dentin (*DSPP*) or enamel-related (*AMTN*, *AMBN*, *ENAM*, *AMELX*, and *MMP20*) patterns of gene expression. Of the hundreds of genes associated with tooth development, most are pleiotropic and perform essential functions outside of tooth development (*15*). By contrast, the aforementioned six genes have been hypothesized to be tooth-specific based on mutagenesis studies in mice, inactivation of these genes in one or more edentulous/enamelless vertebrate species, and natural genetic variation in humans that causes nonsyndromic amelogenesis imperfecta, dentinogenesis imperfecta, and dentin dysplasia (*11*–*13*, *16*–*20*) (table S1). At the same time, other evidence indicates that these six genes are pleiotropic and are expressed outside of tooth development (*21*–*23*) (table S1). In addition to querying bird genomes, we also queried the genomes of additional edentulous/enamelless vertebrates, including three cryptodiran turtles [*Chrysemys picta bellii* (Western painted turtle), *Chelonia mydas* (green sea turtle), and *Pelodiscus sinensis* (Chinese soft-shelled turtle)] and four mammals [*Dasypus novemcinctus* (nine-banded armadillo), *Choloepus hoffmanni* (Hoffmann's two-toed sloth), *Orycteropus afer* (aardvark), and *Manis pentadactyla* (Chinese pangolin)] for signatures of molecular decay in these genes. As a control, we queried the genomes of additional mammalian taxa with enamel-capped teeth [*Canis lupus familiaris* (domestic dog), *Loxodonta africana* (African bush elephant), and *Physeter macrocephalus* (giant sperm whale)].

AMTN, *AMBN*, *ENAM*, *AMELX*, *MMP20*, and *DSPP* were recovered from the genomes of living birds, albeit with numerous inactivating mutations that render these genes nonfunctional (figs. S1 to S8 and tables S2 and S5 to S10). However, the synteny of these six tooth-related genes is strongly conserved across species. The only exceptions occur in *Meleagris gallopavo* (turkey), where *ENAM* exon 9 is found on both chromosome 1 and Z (*13*), and in the *Pterocles guturalis* (sandgrouse), where there appear to be multiple copies of exon 5 (four copies), exon 6 (three copies), and exon 10 (two copies) of *MMP20* scattered across five contigs (fig. S7). Similar to other vertebrate genomes, including human and crocodile, we found that *AMBN* is located between *AMTN* and *ENAM* in the secretory calcium-binding phosphoprotein (SCPP) gene cluster (*24*) (Fig. 2), *AMELX* is found within the *ARHGAP6* gene (Fig. 2), *DSPP* is located within the SIBLING gene cluster between *SPARCL1* and *DMP1* (*24*) (Fig. 2), and *MMP20* is clustered with other matrix metalloproteinase proteins between *MMP27* and *MMP7* (*25*) (Fig. 2 and fig. S8). These synteny results provide unequivocal evidence that the avian pseudogenes for *AMBN*, *AMELX*, *AMTN*, *DSPP*, *ENAM*, and *MMP20* are strictly orthologous with functional copies of these genes in vertebrates with enamel-capped teeth.

Loss of dentin and enamel genes

The remnants of *DSPP*, *AMBN*, *AMELX*, *AMTN*, *ENAM*, and *MMP20* genes in all bird genomes are characterized by exonic deletions and/or other inactivating mutations, including exon deletions, frameshift mutations, stop codons, and

[1]Department of Biology and Molecular Biology, Montclair State University, Montclair, NJ 07043, USA. [2]China National GeneBank, Beijing Genomics Institute-Shenzhen, Shenzhen, 518083, China. [3]Centre for Social Evolution, Department of Biology, Universitetsparken 15, University of Copenhagen, DK-2100 Copenhagen, Denmark. [4]Center for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen Øster Voldgade 5-7, 1350 Copenhagen, Denmark. [5]Trace and Environmental DNA Laboratory, Department of Environment and Agriculture, Curtin University, Perth, Western Australia 6102, Australia. [6]Department of Neurobiology, Howard Hughes Medical Institute and Duke University Medical Center, Durham, NC 27710, USA. [7]Department of Biology, University of California, Riverside, CA 92521, USA.
*Corresponding author. E-mail: meredithr@mail.montclair.edu (R.W.M.); mark.springer@ucr.edu (M.S.S.)

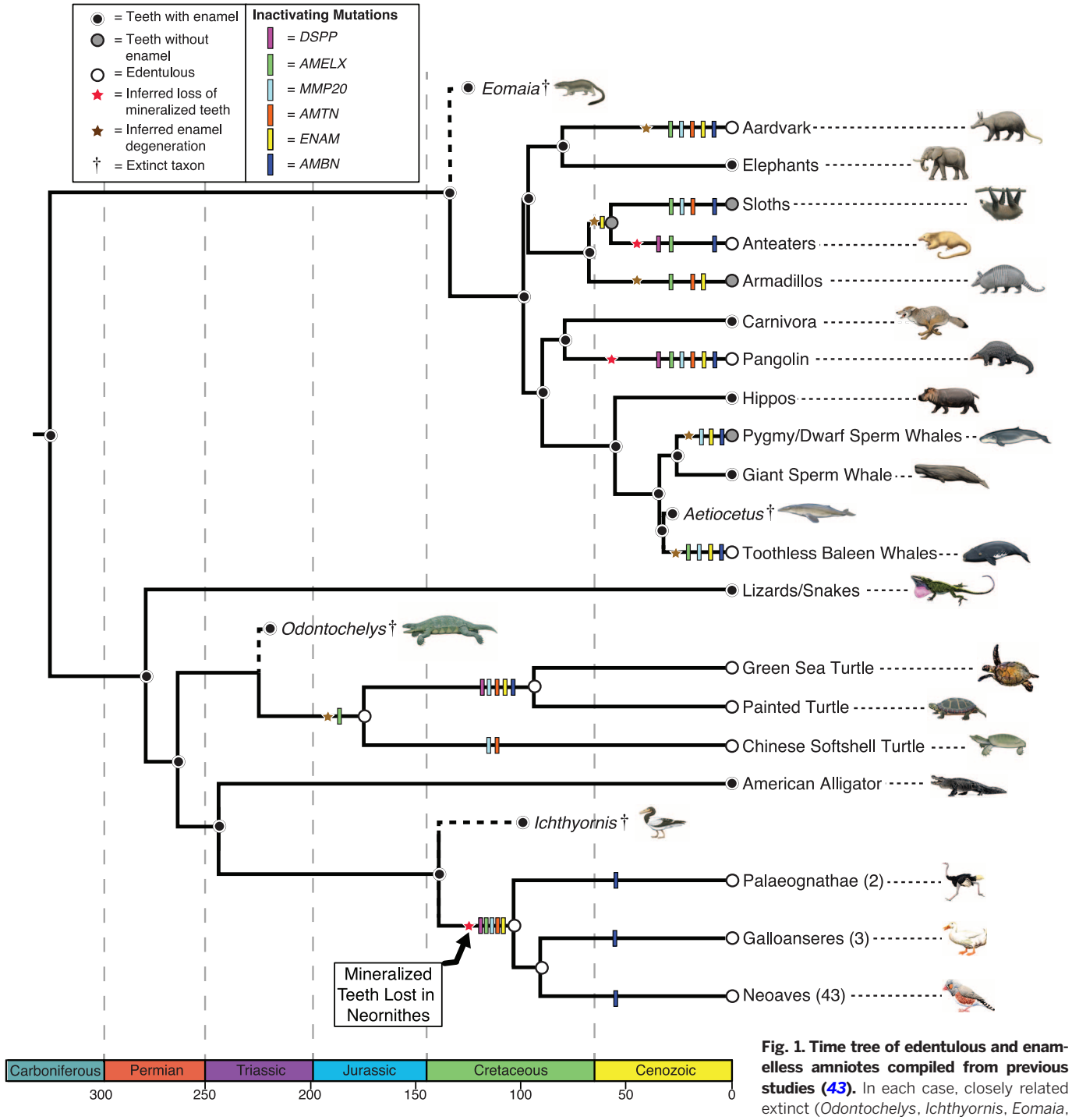


Fig. 1. Time tree of edentulous and enamelless amniotes compiled from previous studies (*43*). In each case, closely related extinct (*Odontochelys*, *Ichthyornis*, *Eomaia*, and *Aetiocetus*) and extant dentate taxa are also shown. Estimates for the timing of enamel loss in mammals are from fossils, phylogenetics, molecular clocks, frameshift mutations, and/or ratio of nonsynonymous to synonymous substitutions (dN/dS) (*4*, *12*). Convergent enamel loss within Xenarthra is inferred for Pilosa (sloths and anteaters) and multiple armadillo lineages (*4*). Information on the enam- and dentin-related genes for pygmy/dwarf sperm whales, anteaters, and baleen whales are from previous studies (*4*, *11*, *12*, *16*, *44*). Hypothesized loss of mineralized teeth on the stem neornithine branch is assumed on the basis of indel substitution rates (*27*, *43*). Numbers in brackets indicate the number of bird genomes analyzed for that particular group. Ghost branches are arbitrarily shown as representing ~5 million years, except for *Ichthyornis*, where the ghost branch was extended to accommodate the depiction of inactivating mutations on the stem Neornithes branch.

splice-site mutations (figs. S1 to S8; tables S2 and S5 to S10; and nexus files 21 to 26). Importantly, all 48 neornithine species share inactivating mutations in both dentin-related (*DSPP*) and enamel-related genes (*ENAM*, *AMELX*, *AMTN*, and *MMP20*) (Fig. 1, Table 1, and tables S5 and S7 to S10). This pattern of shared inactivating mutations indicates that the genetic machinery for tooth formation, including both dentin and enamel, was lost on the branch leading to all extant birds. Additional frameshift mutations in these genes have continued to accumulate in crown lineages and are diagnostic of Palaeognathae, Neognathae, and more nested lineages within these clades (Table 1). The remnants of

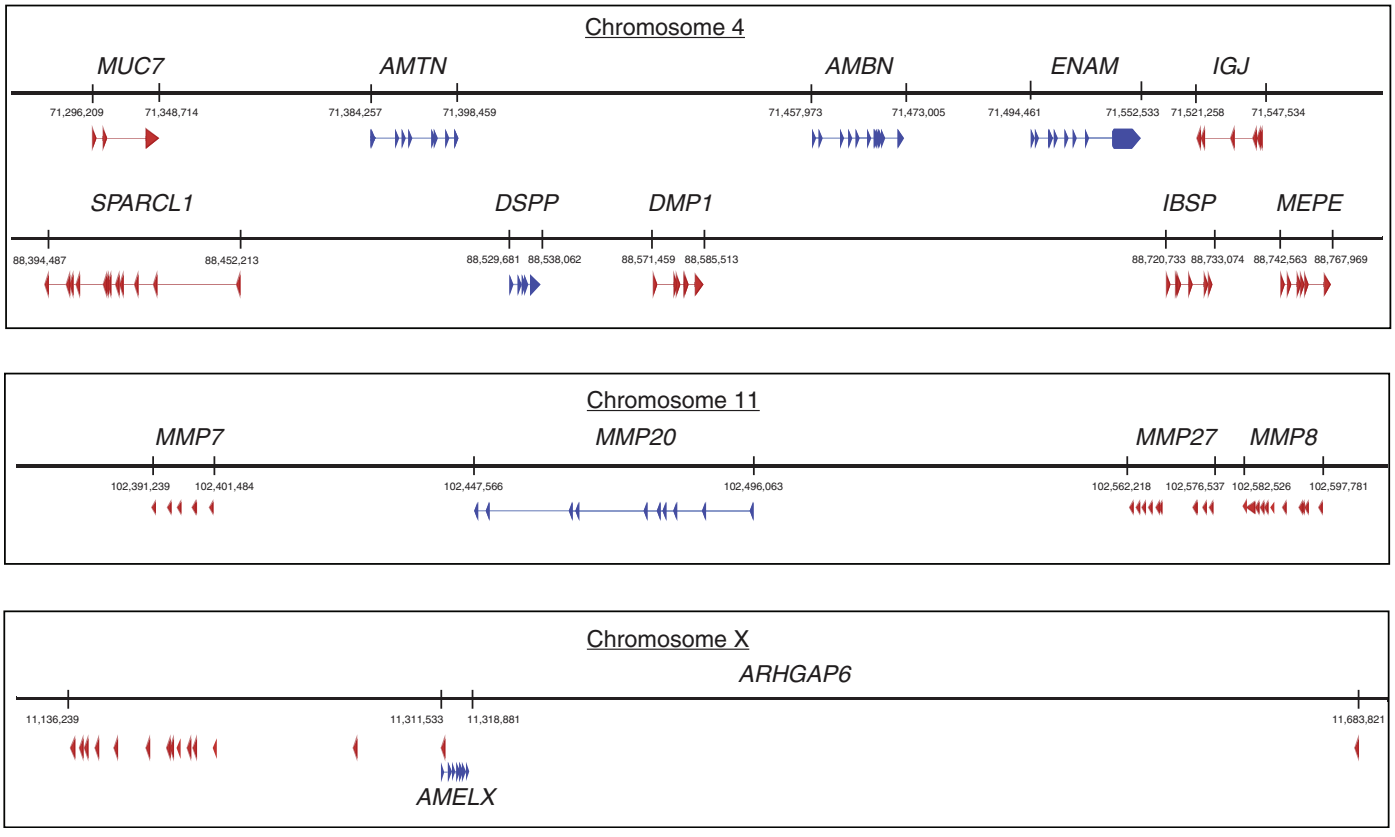


Fig. 2. Schematic representation of enam- and dentin-related genes (blue) and flanking genes (red) from their positions in *Homo* (Ensembl 67), one of the most accurately annotated vertebrate genomes. Exons are indicated by arrows. Details of likely inactivating mutations identified in all enamelless and edentulous species are provided in tables S4 to S21.

Table 1. Total number of shared inactivating mutations in enamel- and dentin-related genes within birds.

	ENAM	*AMBN*	*AMTN*	*AMELX*	*MMP20*	*DSPP*
Neornithes	4	0	1	1	1	4
Palaeognathae	13	3	1	0	2	30
Neognathae	2	0	2	1	2	0
Galloanserae	1	1	1	2	0	1
Neoaves	4	0	2	0	3	4

these genes in extant birds are in advanced states of molecular decay, and we cataloged more than 1300 independent exon deletions and frameshift mutations in these six genes in crown Aves (tables S2 and S5 to S10).

Most of the *DSPP* gene has been lost from the genomes of neognath taxa, whereas relicts of this gene can be found in the two palaeognaths (ostrich and tinamou). Of these, *Struthio camelus* (ostrich) *DSPP* is more complete and retains portions of exon 1 (noncoding) and all five protein-coding exons (2 to 6) (figs. S5 and S6). *Tinamus guttatus* (tinamous), in turn, retains fragments of exons 5 and 6, as well as introns 1, 2, 4, and 5. Sire *et al.* (*26*) reported the presence of *DSPP* exon 2 in *Gallus gallus* (chicken), but we did not find any remnants of exonic regions of *DSPP* in chicken or other members of Galloanserae. Neoaves species retain ~1.5 kb of intron 4, fragments of exon 5 that preserve frameshift mutations that are shared with paleognaths, and portions of the noncoding region of exon 6. The region separating *SPARCL1* and *DMP1* is truncated in birds (5 to 37 kb) relative to American alligator (45 kb). The diminutive size of this region was noted in *Gallus* (*26*), and this pattern is now documented in all newly sequenced bird genomes.

Remnants of *AMBN*, *AMELX*, *AMTN*, *ENAM*, and *MMP20* are more complete than *DSPP* (figs. S1 to S8 and tables S5 to S10), but in some cases whole exons or multiple-exon blocks have been excised from one or more genomes. For example, a block of *ENAM* containing exons 2 to 8 has been deleted in *Calypte anna* (Anna's hummingbird) (fig. S3 and table S2), and exon 1 of *MMP20* is absent from the genomes of living birds (figs. S7 and S8 and tables S2 and S10). *MMP20* is closely syntenic with other matrix metalloproteinase (*MMP*) genes in gnathostomes with sequenced genomes (*25*). In American alligator, *MMP20* is located between *MMP27* and *MMP7*, and the region separating *MMP27* and *MMP20* is ~41 kb (fig. S8). By contrast, this region is reduced by almost 30 kb across a range of diverse bird genomes, including both Palaeognathae and Neognathae, which suggests that exon 1 was excised from the genome in the common ancestor of extant birds (fig. S8). In addition to the loss of exon 1, exon 3 is present only in *Anas platyrhynchos domestica* (domestic duck) (figs. S7 and S8 and table S2), which suggests multiple independent losses of this exon in birds.

Inactivating mutations are also present in these dentin- and enamel-related genes in the three turtle species with genome sequences, including inactivating mutations in *AMELX* shared by the three turtle species we examined (figs. S11 and S12 and tables S17 to S22). However, the genomic record of the six tooth-related genes in turtles is largely incomplete, and additional shared inactivating mutations may have been lost by subsequent, more inclusive deletions. For example, exons 5 to 7 of *ENAM* are missing from all three turtle species (fig. S11), but one or more of these exons may have been deleted in the common ancestor of these taxa followed by subsequent overprinting (e.g., additional deletions) in *Chrysemys* and *Pelodiscus* (fig. S11 and table S3). These difficulties in reconstructing frameshift mutations in tooth-related genes in

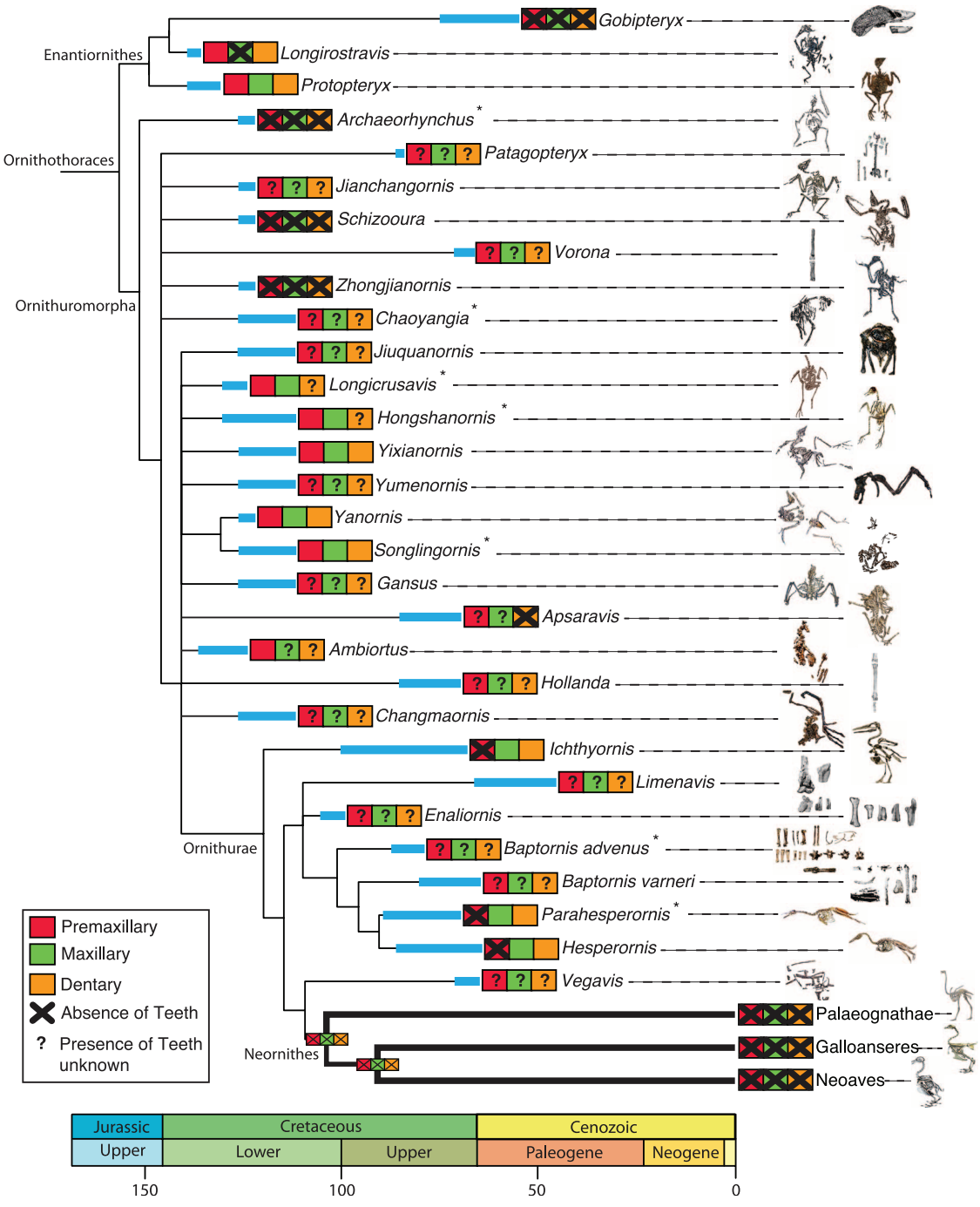


Fig. 3. Time tree of representative ornithurans, indicating multiple iterations of tooth loss or reduction. Ornithothoraces phylogeny from Wang et al. (*45*). Divergence times for crown Aves (thick black lines) are calculated on the basis of Jarvis et al. (*14*). Stratigraphic ages for fossil species (thick blue lines) are from the Paleobiology Database (www.paleobiodb.org) and original descriptions when available (*43*). Ghost branches (thin black lines) extend ~5 million years beyond the oldest fossil in each clade or the immediately adjacent crownward node. Distribution or reduction of teeth across Ornithothoraces is from original descriptions/redescriptions (*43*). Asterisks indicate codings for the presence or absence of teeth that differ from original descriptions and/or the most recent review of tooth loss (*5*).

the early evolutionary history of turtles are not unexpected given that tooth loss in Testudines is at least as old as the Upper Jurassic (~145.5 to 161.2 Ma) (Fig. 1) (*13*). The antiquity of tooth loss in turtles may therefore have provided an extended window for tooth gene degradation in this clade (*13*).

Among enamelless mammals with genome sequences (sloth, armadillo, and aardvark), the *DSPP* gene remains intact, whereas *MMP20*, *ENAM*, *AMTN*, *AMELX*, and *AMBN* genes have been inactivated, with the exception of *AMBN* and *MMP20* in armadillo (Fig. 1, figs. S9 and S10, and tables S11 to S16) (*13*). There are no inactivating mutations shared by the two xenarthrans (sloth and armadillo), although sloths share frameshift mutations in the *ENAM* gene with anteaters (*4*). The enamel-related *MMP20*, *ENAM*, *AMTN*, *AMELX*, and *AMBN* genes have also been inactivated in the edentulous Chinese pangolin, and all of the protein-coding exons have been excised from the dentin-related *DSPP* gene in this species (Fig. 1 and tables S11 to S14 and S16) (*13*). The functionality of *DSPP* (pygmy/dwarf sperm whales and baleen whales), *MMP20* (anteaters), *AMTN* (anteaters, pygmy/dwarf sperm whales, and baleen whales), and *AMBN* (anteaters) are currently unknown. We also did not find evidence for molecular decay of tooth genes in the genomes of any of the nonavian toothed sauropsids [*Alligator*, and *Anolis carolinensis* (Carolina anole)] or mammals (elephant, dog, and sperm whale) (Fig. 1) that were included in our study. These results confirm that inactivation of tooth genes has occurred convergently in birds, turtles, and multiple mammalian lineages.

Dating of mutations in genes associated with tooth loss in birds

The occurrence of shared inactivating mutations in both dentin- and enamel- related genes suggests that teeth were lost on the stem neornithine

branch, which extends from 244.2 to 101.6 Ma (*14*). On the basis of the rate at which frameshift mutations accumulate (*27*) in neutrally evolving, tooth-related pseudogenes in crown Neornithes and the avian time tree of the 48 species in (*14*), we estimate that the outer enamel covering of teeth was lost around 116 Ma (95% confidence interval = 105.7 to 128.3 Ma) (see the supplementary materials) in the common ancestor of all modern birds. The remnants of *DSPP* in modern birds are too incomplete to date dentin loss, but if the loss of *DSPP* is a molecular proxy for edentulism then tooth loss in birds occurred no latter than 101.6 Ma on the basis of patterns of shared mutations in *DSPP* in modern birds. The closest relatives of crown Neornithes belong to Ichthyorniformes (e.g., *Guildavis*, *Iaceornis*, and *Ichthyornis*) and Hesperorniformes (e.g., *Baptornis* and *Hesperornis*) (*7*, *28*) and range from the late Cenomanian (*Ichthyornis*, ~93.5 Ma) to within 300,000 years of the Cretaceous-Paleogene (KPg) boundary (*29*) at 66 Ma when modern birds are thought to have undergone a rapid radiation of species (*14*). The presence or absence of teeth cannot be determined for most ichthyorniform and hesperorniform taxa owing to incomplete skeletal remains, but *Ichthyornis* and *Hesperornis* fossils possess both dentary and maxillary teeth (*28*). Our genomic results support the exclusion of these taxa, as well as all dentate bird fossils, from crown Neornithes. Instead, they suggest that members of Ichthyorniformes and Hesperorniformes were dentate contemporaries of edentulous crown birds for most of the Late Cretaceous. Other avian contemporaries of crown Aves in the Late Cretaceous included both toothed and edentulous birds belonging to the now extinct clade Enantiornithes ("opposite birds") (Fig. 2).

Discussion

Previous studies have documented that *DSPP*, *AMBN*, *AMELX*, *AMTN*, *ENAM*, and *MMP20* are expressed outside of tooth development (*21*–*23*) (table S1). Our findings of widespread and convergent pseudogenization of these six genes across diverse edentulous and enamelless vertebrates provides evidence that these genes are tooth-specific with respect to their essential functions that are maintained by natural selection. Viewed from another perspective, these six genes are pleiotropic with the caveat that their pleiotropic roles outside of tooth development are non-obligatory. *DSPP*, which is involved in dentin formation, has been independently inactivated in three lineages of edentulous amniotes (birds, turtles, and Chinese pangolin). Similarly, *AMELX*, *AMBN*, *AMTN*, *ENAM*, and *MMP20* are critical for enamel formation (table S1) and have been independently inactivated in birds, turtles, pangolins, sloths, armadillos (except *MMP20* and *AMBN*), and aardvarks. Anteaters, baleen whales, pygmy/dwarf sperm whales, and narwhals also show evidence of inactivation of enamel-related genes, although sequence coverage for these genes is limited to one or a few partial exons of *AMBN*, *AMELX*, *ENAM*, and *MMP20* (*4*, *11*, *12*). If these genes played critical, essential roles outside of enamel or dentin development, we would instead expect them to be maintained by natural selection in edentulous and enamelless lineages. It is possible that redundancy in the transcriptome may have compensated for their non-tooth–specific functions in edentulous and enamelless lineages. For example, *MMP20*'s primary role is in enamel formation, but it is also expressed in developing dentin and together with MMP2 cleaves DSPP in dentin formation (*30*). The MMP2-MMP20 redundancy in DSPP cleavage (*30*) may have facilitated inactivation of *MMP20* once it was released from performing its function in enamel formation in enamelless taxa that nevertheless retain dentin-based teeth such as aardvark and sloth.

It is also possible that tooth-related genes are incidentally expressed outside of tooth development via gene expression neighborhoods (*31*, *32*). Genes such as *AMBN*, *AMTN*, and *ENAM* occur in the SCPP gene cluster (*24*) that also includes bone-related genes and salivary proteins, and incidental coexpression may be expected to occur in some of these tissues. Expression studies provide support for the expression of these tooth-related genes in other biomineralized tissues (*22*, *33*, *34*) (table S1). Similarly, *MMP20* is located in one of the MMP gene clusters with several other matrix metalloproteinase genes, and its coexpression in nondental tissues (e.g., lung and large intestine) (*21*, *35*, *36*) with other MMPs may be attributed, in part, to gene expression neighborhoods.

The possibility of functional splice site variants also merits attention. Given that multiple splice site variants are associated with Amelogenin, the gene *AMELX* (*37*) may remain intact if inactivating mutations are confined to a subset of *AMELX* exons. However, our data identified widely distributed inactivating mutations in *AMELX* across all of the protein-coding exons of this gene, which suggests that no bird has viable splice site variants (table S9) (*13*).

Edentulism and the development of a complete rhamphotheca are integral features of modern birds. Beak development, in turn, is associated with the expression of an assemblage of beta-keratin genes that differs from the sets of beta-keratin genes associated with other epidermal appendages, including scales, claws, feathers, and the egg tooth (*38*). Expression studies suggest that inactivation of *BMP4* in the oral epithelium of chick embryos leads to early-stage odontogenic arrest (*39*). Moreover, *BMP4* expression levels in beak mesenchyme are correlated with beak morphology (*40*). The inactivation of tooth genes does not provide direct evidence for a complete rhamphotheca, but fossil-based evidence (e.g., *Hesperornis* and *Ichthyornis*) for the replacement of premaxilary teeth by a partial rhamphotheca in the first stage of rostral remodeling suggests that tooth loss and rhamphotheca development are tightly linked, as may be expected if keratinization of epithelial cells impeded tooth formation (*41*).

We postulate, on the basis of fossil and molecular evidence, a two-step scenario whereby tooth loss and rhamphotheca development evolved in concert with each other on the stem neornithine branch. We propose that the first stage included tooth loss on the premaxilla and the development of a partial rhamphotheca on the premaxilla and anterior mandible (*5*, *42*). Direct evidence for this incipient stage in the remaking of the avian rostrum derives from the fossil record (*5*, *42*). Importantly, the evolution of a partial rhamphotheca may have been a prerequisite for complete edentulism so that the latter was adaptively viable (*5*). The second stage involved concurrent rostrocaudal progression of tooth loss and rhamphotheca development from the already modified anterior rostrum (*5*) (Fig. 2). We propose that this progression ultimately resulted in a complete rhamphotheca that effectively replaced the teeth and contributed to the diversification of living birds. (Fig. 3)

REFERENCES AND NOTES

1. J. del Hoyo, A. Elliott, J. Sargatal, Eds. (vols. 1 to 7), and David A Christie, Ed. (vols. 8 to 16), *Handbook of the Birds of the World* (Lynx Edicions, Barcelona, 1992–2013).
2. D. E. Wilson, D. M. Reeder, *Mammal Species of the World* (Johns Hopkins Univ. Press, Baltimore, MD, ed. 3, 2005).
3. Turtle Taxonomy Working Group (P. P. van Dijk, J. B. Iverson, H. B. Shaffer, R. Bour, A. G. J. Rhodin). Turtles of the world, 2012 update: Annotated checklist of taxonomy, synonymy, distribution, and conservation status, in A. G. J. Rhodin, P. C. H. Pritchard, P. P. van Dijk, R. A. Saumure, K. A. Buhlmann, J. B. Iverson, R. A. Mittermeier, Eds., *Conservation Biology of Freshwater Turtles and Tortoises: A Compilation Project of the IUCN/SSC Tortoise and Freshwater Turtle Specialist Group. Chelonian Research Monographs No. 5* (2012), pp. 000.243–000.328; doi:10.3854/crm.5.000.checklist.v5.2012; www.iucn-tftsg.org/cbftt. doi: 10.3854/crm.5.000.checklist.v5.2012
4. R. W. Meredith, J. Gatesy, W. J. Murphy, O. A. Ryder, M. S. Springer, Molecular decay of the tooth gene Enamelin (ENAM) mirrors the loss of enamel in the fossil record of placental mammals. *PLOS Genet.* **5**, e1000634 (2009). doi: 10.1371/journal.pgen.1000634; pmid: 19730686
5. A. Louchart, L. Viriot, From snout to beak: The loss of teeth in birds. *Trends Ecol. Evol.* **26**, 663–673 (2011). doi: 10.1016/j.tree.2011.09.004; pmid: 21978465
6. X. Xu, H. You, K. Du, F. Han, An *Archaeopteryx*-like theropod from China and the origin of Avialae. *Nature* **475**, 465–470 (2011). doi: 10.1038/nature10288; pmid: 21796204
7. J. K. O'Connor, Z. Zhou, A redescription of *Chaoyangia beishanensis* (Aves) and a comprehensive phylogeny of Mesozoic birds. *J. Syst. Palaeontology* **11**, 889–906 (2013). doi: 10.1080/14772019.2012.690455
8. J. Cracraft, The origin and early diversification of birds. *Paleobiology* **12**, 383–399 (1986).
9. B. C. Livezey, R. L. Zusi, Higher-order phylogeny of modern birds (Theropoda, Aves: Neornithes) based on comparative anatomy. II. Analysis and discussion. *Zool. J. Linn. Soc.* **149**, 1–95 (2007). doi: 10.1111/j.1096-3642.2006.00293.x; pmid: 18784798
10. J. Cracraft, Phylogenetic relationships and monophyly of loons, grebes, and hesperornithiforms, with comments on the early history of birds. *Syst. Zool.* **31**, 35–56 (1982). doi: 10.2307/2413412
11. T. A. Demeré, M. R. McGowen, A. Berta, J. Gatesy, Morphological and molecular evidence for a stepwise evolutionary transition from teeth to baleen in mysticete whales. *Syst. Biol.* **57**, 15–37 (2008). doi: 10.1080/10635150701884632; pmid: 18266181
12. R. W. Meredith, J. Gatesy, J. Cheng, M. S. Springer, Pseudogenization of the tooth gene enamelysin (*MMP20*) in the common ancestor of extant baleen whales. *Proc. Biol. Sci.* **278**, 993–1002 (2011). doi: 10.1098/rspb.2010.1280; pmid: 20861053
13. R. W. Meredith, J. Gatesy, M. S. Springer, Molecular decay of enamel matrix protein genes in turtles and other edentulous amniotes. *BMC Evol. Biol.* **13**, 20 (2013). doi: 10.1186/1471-2148-13-20; pmid: 23342979

14. E. D. Jarvis *et al*., Whole genome analyses resolve the early branches in the tree of life of modern birds. *Science* **346**, 1320–1331 (2014).
15. I. Thesleff, The genetic basis of tooth development and dental defects. *Am. J. Med. Genet.* **140A**, 2530–2535 (2006). doi: 10.1002/ajmg.a.31360; pmid: 16838332
16. D. A. McKnight, L. W. Fisher, Molecular evolution of dentin phosphoprotein among toothed and toothless animals. *BMC Evol. Biol.* **9**, 299 (2009). doi: 10.1186/1471-2148-9-299; pmid: 20030824
17. B. Gasse, J. Silvent, J.-Y. Sire, Evolutionary analysis suggests that AMTN is enamel-specific and a candidate for AI. *J. Dent. Res.* **91**, 1085–1089 (2012). doi: 10.1177/0022034512460551; pmid: 22968158
18. J. D. Bartlett *et al*., Protein-protein interactions of the developing enamel matrix. *Curr. Top. Dev. Biol.* **74**, 57–115 (2006). doi: 10.1016/S0070-2153(06)74003-0; pmid: 16860665
19. J. Hatakeyama *et al*., Synergistic roles of amelogenin and ameloblastin. *J. Dent. Res.* **88**, 318–322 (2009). doi: 10.1177/0022034509334749; pmid: 19407150
20. J. C.-C. Hu *et al*., Enamel defects and ameloblast-specific expression in Enam knock-out/lacz knock-in mice. *J. Biol. Chem.* **283**, 10858–10871 (2008). doi: 10.1074/jbc.M710565200; pmid: 18252720
21. K. J. Greenlee, Z. Werb, F. Kheradmand, Matrix metalloproteinases in lung: Multiple, multifarious, and multifaceted. *Physiol. Rev.* **87**, 69–98 (2007). doi: 10.1152/physrev.00022.2006; pmid: 17237343
22. D. Deutsch *et al*., Amelogenin, a major structural protein in mineralizing enamel, is also expressed in soft tissues: Brain and cells of the hematopoietic system. *Eur. J. Oral Sci.* **114** (suppl. 1), 183–189, discussion 201–202, 381 (2006). doi: 10.1111/j.1600-0722.2006.00301.x; pmid: 16674683
23. J. Vymětal, I. Slabý, A. Spahr, J. Vondrásek, S. P. Lyngstadaas, Bioinformatic analysis and molecular modelling of human ameloblastin suggest a two-domain intrinsically unstructured calcium-binding protein. *Eur. J. Oral Sci.* **116**, 124–134 (2008). doi: 10.1111/j.1600-0722.2008.00526.x; pmid: 18353005
24. K. Kawasaki, K. M. Weiss, SCPP gene evolution and the dental mineralization continuum. *J. Dent. Res.* **87**, 520–531 (2008). doi: 10.1177/154405910808700608; pmid: 18502959
25. K. Kawasaki, T. Suzuki, Molecular evolution of matrix metalloproteinase 20. *Eur. J. Oral Sci.* **119** (suppl. 1), 247–253 (2011). doi: 10.1111/j.1600-0722.2011.00898.x; pmid: 22243253
26. J.-Y. Sire, S. C. Delgado, M. Girondot, Hen's teeth with enamel cap: From dream to impossibility. *BMC Evol. Biol.* **8**, 246 (2008). doi: 10.1186/1471-2148-8-246; pmid: 18775069
27. M. S. Springer, A. Burk, J. R. Kavanagh, V. G. Waddell, M. J. Stanhope, The interphotoreceptor retinoid binding protein gene in therian mammals: Implications for higher level relationships and evidence for loss of function in the marsupial mole. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 13754–13759 (1997). doi: 10.1073/pnas.94.25.13754; pmid: 9391099
28. L. M. Chiappe, L. M. Witmer, Mesozoic birds: above the heads of dinosaurs. (University of California Press, Berkeley, Calif., 2002).
29. N. R. Longrich, T. Tokaryk, D. J. Field, Mass extinction of birds at the Cretaceous-Paleogene (K-Pg) boundary. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 15253–15257 (2011). doi: 10.1073/pnas.1110395108; pmid: 21914849
30. Y. Yamakoshi *et al*., Dentin sialophosphoprotein is processed by MMP-2 and MMP-20 *in vitro* and *in vivo*. *J. Biol. Chem.* **281**, 38235–38243 (2006). doi: 10.1074/jbc.M607767200; pmid: 17046814
31. B. Oliver, M. Parisi, D. Clark, Gene expression neighborhoods. *J. Biol.* **1**, 4 (2002). doi: 10.1186/1475-4924-1-4; pmid: 12144705
32. P. T. Spellman, G. M. Rubin, Evidence for large domains of similarly expressed genes in the Drosophila genome. *J. Biol.* **1**, 5 (2002). doi: 10.1186/1475-4924-1-5; pmid: 12144710
33. A. Haze *et al*., Amelogenin expression in long bone and cartilage cells and in bone marrow progenitor cells. *Anat. Rec.* **290**, 455–460 (2007). doi: 10.1002/ar.20520; pmid: 17393535
34. P. Atsawasuwan *et al*., Expression and function of enamel-related gene products in calvarial development. *J. Dent. Res.* **92**, 622–628 (2013). doi: 10.1177/0022034513487906; pmid: 23625374
35. J. Ryu *et al*., Differential expression of matrix metalloproteinases and their inhibitors in human and mouse lung development. *Thromb. Haemost.* **94**, 175–183 (2005). pmid: 16113801
36. B. E. Turk *et al*., MMP-20 is predominately a tooth-specific enzyme with a deep catalytic pocket that hydrolyzes type V collagen. *Biochemistry* **45**, 3863–3874 (2006). doi: 10.1021/bi052252o; pmid: 16548514
37. A. K. Mathur, P. D. Polly, The evolution of enamel microstructure: How important is amelogenin? *J. Mamm. Evol.* **7**, 23–42 (2000). doi: 10.1023/A:1009413632741
38. M. J. Greenwold, R. H. Sawyer, Genomic organization and molecular phylogenies of the beta (beta) keratin multigene family in the chicken (Gallus gallus) and zebra finch (Taeniopygia guttata): Implications for feather evolution. *BMC Evol. Biol.* **10**, 148 (2010). doi: 10.1186/1471-2148-10-148; pmid: 20482795
39. Y. Chen *et al*., Conservation of early odontogenic signaling pathways in Aves. *Proc. Natl. Acad. Sci. U.S.A.* **97**, 10044–10049 (2000). doi: 10.1073/pnas.160245097; pmid: 10954731
40. A. Abzhanov, M. Protas, B. R. Grant, P. R. Grant, C. J. Tabin, *Bmp4* and morphological variation of beaks in Darwin's finches. *Science* **305**, 1462–1465 (2004). doi: 10.1126/science.1098095; pmid: 15353802
41. M. P. Harris, S. M. Hasso, M. W. J. Ferguson, J. F. Fallon, The development of archosaurian first-generation teeth in a chicken mutant. *Curr. Biol.* **16**, 371–377 (2006). doi: 10.1016/j.cub.2005.12.047; pmid: 16488870
42. T. L. Hieronymus, L. M. Witmer, Homology and evolution of avian compound rhamphothecae. *Auk* **127**, 590–604 (2010). doi: 10.1525/auk.2010.09122
43. Supplementary materials are available on *Science* Online.
44. S. Delgado, N. Vidal, G. Veron, J.-Y. Sire, Amelogenin, the major protein of tooth enamel: A new phylogenetic marker for ordinal mammal relationships. *Mol. Phylogenet. Evol.* **47**, 865–869 (2008). doi: 10.1016/j.ympev.2008.01.025; pmid: 18346915
45. Y.-M. Wang, J. K. O'Connor, D.-Q. Li, H.-L. You, Previously unrecognized ornithuromorph bird diversity in the Early Cretaceous Changma Basin, Gansu Province, northwestern China. *PLOS ONE* **8**, e77693 (2013). doi: 10.1371/journal.pone.0077693; pmid: 24147058

ACKNOWLEDGMENTS

This work was supported by Montclair State University FY2014 Separately Budgeted Research Internal Reward and Startup Funds to R.W.M. and NSF (EF0629860) to M.S.S. Paintings are by J. Fjeldså (birds), R.W.M. (green sea turtle, Caroline anole, alligator, and Fig. 3), and C. Buell (all others). Comparative genomic alignments are in the supplementary materials (nexus files 1 to 38).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/346/6215/1254390/suppl/DC1
Supplementary Text
Figs. S1 to S12
Tables S1 to S22
References (*46*–*111*)

4 April 2014; accepted 10 September 2014
10.1126/science.1254390

RESEARCH ARTICLE SUMMARY

AVIAN GENOMICS

Statistical binning enables an accurate coalescent-based estimation of the avian tree

Siavash Mirarab, Md. Shamsuzzoha Bayzid, Bastien Boussau, Tandy Warnow*

INTRODUCTION: Reconstructing species trees for rapid radiations, as in the early diversification of birds, is complicated by biological processes such as incomplete lineage sorting (ILS) that can cause different parts of the genome to have different evolutionary histories. Statistical methods, based on the multispecies coalescent model and that combine gene trees, can be highly accurate even in the presence of massive ILS; however, these methods can produce species trees that are topologically far from the species tree when estimated gene trees have error. We have developed a statistical binning technique to address gene tree estimation error and have explored its use in genome-scale species tree estimation with MP-EST, a popular coalescent-based species tree estimation method.

ON OUR WEB SITE

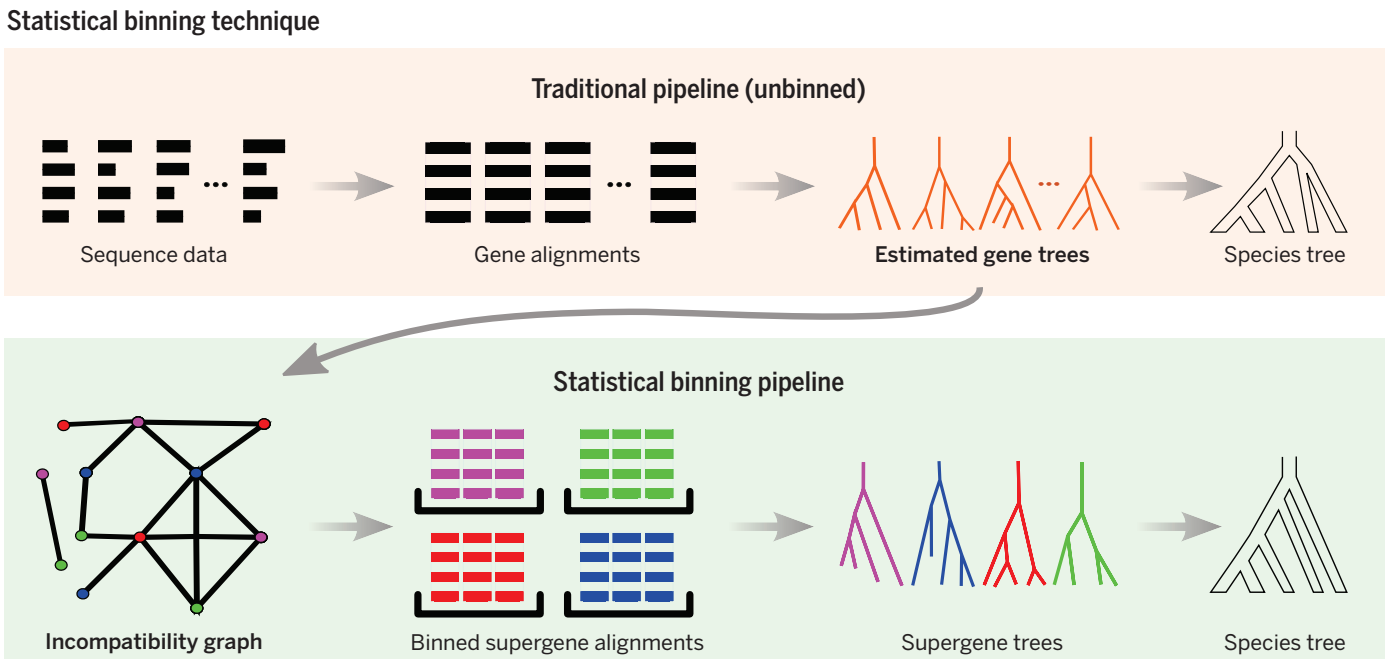
Read the full article at http://dx.doi.org/10.1126/science.1250463

RATIONALE: In statistical binning, phylogenetic trees on different genes are estimated and then placed into bins, so that the differences between trees in the same bin can be explained by estimation error (see the figure). A new tree is then estimated for each bin by applying maximum likelihood to a concatenated alignment of the multiple sequence alignments of its genes, and a species tree is estimated using a coalescent-based species tree method from these supergene trees.

RESULTS: Under realistic conditions in our simulation study, statistical binning reduced the topological error of species trees estimated using MP-EST and enabled a coalescent-based analysis that was more accurate than concatenation even when gene tree estimation error was relatively high. Statistical binning also reduced the error in gene tree topology and species tree branch length estimation, especially when the phylogenetic signal in gene sequence alignments was low. Species trees estimated using MP-EST with statistical binning on four biological data sets showed increased concordance with the biological literature. When MP-EST was used to analyze 14,446 gene trees in the avian phylogenomics project, it produced a species tree that was discordant with the concatenation analysis and conflicted with prior literature. However, the statistical binning analysis produced a tree that was highly congruent with the concatenation analysis and was consistent with the prior scientific literature.

CONCLUSIONS: Statistical binning reduces the error in species tree topology and branch length estimation because it reduces gene tree estimation error. These improvements are greatest when gene trees have reduced bootstrap support, which was the case for the avian phylogenomics project. Because using unbinned gene trees can result in overestimation of ILS, statistical binning may be helpful in providing more accurate estimations of ILS levels in biological data sets. Thus, statistical binning enables highly accurate species tree estimations, even on genome-scale data sets. ■

The list of author affiliations is available in the full article online.
*Corresponding author. E-mail: warnow@illinois.edu
Cite this article as S. Mirarab *et al.*, *Science* 346, 1250463 (2014). DOI: 10.1126/science.1250463



The statistical binning pipeline for estimating species trees from gene trees. Loci are grouped into bins based on a statistical test for combinabilty, before estimating gene trees.

Published by AAAS

Downloaded from www.sciencemag.org on December 14, 2014

RESEARCH ARTICLE

Statistical binning enables an accurate coalescent-based estimation of the avian tree

Siavash Mirarab,[1] Md. Shamsuzzoha Bayzid,[1] Bastien Boussau,[2] Tandy Warnow[1,3]*

Gene tree incongruence arising from incomplete lineage sorting (ILS) can reduce the accuracy of concatenation-based estimations of species trees. Although coalescent-based species tree estimation methods can have good accuracy in the presence of ILS, they are sensitive to gene tree estimation error. We propose a pipeline that uses bootstrapping to evaluate whether two genes are likely to have the same tree, then it groups genes into sets using a graph-theoretic optimization and estimates a tree on each subset using concatenation, and finally produces an estimated species tree from these trees using the preferred coalescent-based method. Statistical binning improves the accuracy of MP-EST, a popular coalescent-based method, and we use it to produce the first genome-scale coalescent-based avian tree of life.

Species trees provide a basis for understanding how life evolved on earth, as well as having applications to comparative genomics, orthology detection, protein function inference, and biodiversity analysis. Estimations of species trees are typically built with multiple loci (sometimes complete genes—but not always), in some cases, from throughout the genome. One advantage of such a phylogenomic approach is that it enables more data to be used in tree estimation (*1*). However, there is increasing evidence that loci can have conflicting evolutionary histories (so that their phylogenetic trees are topologically different) because of many biological causes, including incomplete lineage sorting (ILS) (*2*, *3*), a process that is especially common in rapid radiations, characterized by a succession of short branches in the phylogenetic tree, such as is believed to have occurred in the avian and mammalian evolutionary lineages (*4*–*9*). However, the standard phylogenetic estimation technique of concatenation, which concatenates alignments for individual loci into a combined data set called a supermatrix and then estimates the species tree from the supermatrix, can return incorrect species trees with high confidence in the presence of substantial ILS (*10*–*13*).

For this reason, many methods have been developed to estimate species trees that can be accurate even with high levels of ILS (*4*, *8*, *12*–*19*). For example, coestimation methods produce estimated gene trees and species trees directly from sequence alignments (*14*, *15*, *18*), and summary methods operate by combining estimated gene trees into a species tree. (Not all the loci in a phylogenomic analysis may be complete genes, e.g., some may contain only the exons or only the introns of some gene, and some may not be based on genes at all. However, to be consistent with other literature on the subject (*2*, *10*), we refer to phylogenetic trees on genomic loci as gene trees.) Some of these summary methods are created on the basis of the multispecies coalescent model (*20*) and are statistically consistent under that model (*12*, *13*), which means they will reconstruct the true species tree with high probability given a sufficiently large number of estimated gene trees that are error-free (*12*, *13*, *17*, *18*). A new type of coalescent-based method estimates the species tree directly from unlinked markers without also estimating gene trees (*21*–*23*).

However, the performance of coalescent-based methods has been mixed. Coestimation methods can have excellent accuracy but are too computationally intensive to use on data sets with hundreds of genes (*24*). The methods that estimate species trees directly from the sequence data without also estimating gene trees are not as computationally intensive as the coestimation methods but are much less well understood because they have only recently been developed. Furthermore, some [like SNAPP (*21*)] can only be used with biallelic markers and so are not suitable for estimating species trees for large data sets with deep divergences where biallelic markers are rare. Summary methods are by far the most frequently used method for species tree estimation and have produced good results on some biological data sets (*14*, *15*); however, for other data sets, the summary methods have not been able to produce highly supported trees (*25*), even with a large quantity of data (*26*). Simulation studies show that species trees estimated with summary methods can be less accurate than species trees estimated with concatenation, even in the presence of substantial ILS (*11*, *27*, *28*). A main reason for this disparity in performance is poor phylogenetic signal (e.g., because of short sequence lengths) in individual genes, which is a potential problem for coalescent-based summary methods (*28*, *29*). Moreover, many realistic biological conditions (including short branches in gene trees) make completely accurate gene tree estimation from limited sequence data highly unlikely (*30*).

Phylogenomic analyses can utilize very large numbers of genomic loci to estimate the species tree, but genome-scale data sets can contain loci that have reduced phylogenetic signal so that their estimated gene trees have reduced bootstrap support (BS) (*31*). Although it is not known how summary methods are affected when only some of the loci have low signal, coalescent-based summary methods have reduced accuracy on data sets where all the gene sequence alignments are short (*28*).

This challenge confronted the avian phylogenomics project (*31*), where a species tree estimated with a concatenated maximum-likelihood analysis on 14,446 loci had a succession of short branches suggestive of a radiation, and the tree also conflicted with estimated gene trees. Furthermore, most loci had low phylogenetic signal, which resulted in average BS of only 32% for the bifurcating maximum-likelihood trees estimated on these loci (fig. S1). Although much of the distance between estimated gene trees and the estimated species tree was related to the low support branches, even after collapsing low support branches, there was still substantial conflict among the gene trees (fig. S15 and supplementary text), suggestive of ILS. Thus, not only is there gene tree conflict (reducing the accuracy of concatenation) but the gene trees were generally poorly estimated (reducing the accuracy of summary methods).

Constructing phylogenies from genes with low phylogenetic signal is challenging, even if ILS is not an issue, and several approaches for selecting loci for use within a concatenated analysis have been suggested (*32*). However, restricting loci is problematic for statistically consistent coalescent-based summary methods, because the conditions under which they are guaranteed to be accurate (with high probability) require a large enough random sample of true gene trees; removing loci can violate this condition and potentially bias the analysis.

Statistical binning technique

A phylogenomic pipeline that uses a coalescent-based summary method begins with sequence alignments on different loci, estimates gene trees on each locus, and then combines the estimated gene trees into an estimated species tree using the summary method. The statistical binning step modifies the pipeline by using a binning technique to produce a different set of estimated gene trees that can be used with the summary method. We call the use of binning with a given summary method the binned version of the summary method.

How the statistical binning pipeline operates, given an input set of loci with their estimated

[1]Department of Computer Science, University of Texas at Austin, Austin, TX 78712, USA. [2]Laboratoire de Biométrie et Biologie Evolutive, CNRS, UMR5558, Université Lyon 1, 69622, Villeurbanne, France. [3]Department of Bioengineering and Computer Science, University of Illinois Urbana-Champaign, Champaign, IL 61820, USA.
*Corresponding author. E-mail: warnow@illinois.edu

sequence alignments and trees is shown in Fig. 1. We use BS values on branches of the estimated gene trees to partition the set of loci into bins of roughly equal sizes, so that each bin consists of a set of loci where differences in the estimated gene trees can be explained by gene tree estimation error. We concatenate the alignments of loci in each bin into a large alignment (called a supergene alignment) and compute trees on each supergene alignment using maximum likelihood (*33*); this produces a set of trees (called supergene trees), with one supergene tree for each bin. We then construct a species tree from the set of supergene trees using the desired summary method. Thus, the difference between the unbinned and binned versions of a summary method is the set of trees it uses to compute the species tree: The unbinned summary method uses the original set of gene trees, and the binned summary method uses the set of supergene trees.

Evaluation

We used biological and simulated data sets (*34*) to evaluate species trees estimated by using binned and unbinned summary methods, as well as concatenation using maximum likelihood under the GTR+Γ model, computed by randomized accelerated maximum likelihood (RAxML) (*33*). Three summary methods—the greedy consensus, matrix representation with parsimony (MRP) (*35*), and maximum pseudo-likelihood estimation of species trees (MP-EST) (*13*)—were applied to simulated avian and mammalian data sets by using the site-only multilocus bootstrapping procedure (*34*) [see (*36*) for a discussion of more elaborate approaches for bootstrapping multilocus data sets]. We chose MP-EST because it is statistically consistent under the multispecies coalescent model, has been used in several studies (*6*, *37*–*39*), and had better accuracy than other summary methods in some studies (*13*). However, the greedy consensus is inconsistent under the multispecies coalescent (*40*), and MRP and concatenation may also be inconsistent (*11*).

We generated simulated data sets (*34*) from two model species trees: one based on the avian phylogenomics project data set with 48 species and 14,446 loci (*31*), and one based on a mammalian data set with 37 species and 447 loci studied in (*6*). The default model conditions center on the average gene tree BS and ILS levels of these biological data sets, and we varied the model parameters to produce lower and higher ILS levels and estimated gene trees with varying BS to understand the impact of binning under a wide range of conditions. Each model species tree was computed by running MP-EST on the biological data, and we modified the branch lengths on the model species trees to produce other model conditions with different amounts of ILS (the 2× condition has doubled branch lengths and so reduces ILS, and the 0.5× condition has halved branch lengths and so increases ILS). We simulated gene trees within the model species trees (table S1) on the basis of the multispecies coalescent model (*20*). We evolved sequences of different lengths down the gene trees and used RAxML (*33*) with 200 bootstrap replicates to estimate gene trees with branch support on these sequence alignments.

The avian biological gene trees have very low average BS. Of the three types of genomic markers—exons, introns, and UCEs (ultraconserved markers) analyzed—the exons have the least signal (average BS 24%), the introns have the most (average BS 48%), and the UCEs are intermediate in support (average BS 39%). The longest introns (with at least 10,000 bp) have the highest average BS (59%) but represent a very small fraction of the total set of gene trees examined (only 638 of 14,446).

We modeled conditions that resembled the avian exons-only, UCEs-only, introns-only, and long introns–only data sets (fig. S1), with respect to their average BS values, and refer to these different model conditions by the partition type (below). The simulated mammalian data sets exhibit support levels of 63 and 79%, bracketing the 71% average BS values in the biological data. We varied the number of genes from 200 to 2000 for the avian data set and from 200 to 800 for the mammalian data set. Finally, we created mixed-model conditions, one with 14,350 genes for the avian simulation experiment and the other with 400 genes for the mammalian simulation experiment, to closely approximate the biological data sets in terms of the number of loci and average BS. Overall, the avian simulated data sets have higher levels of ILS and lower BS values than the mammalian data sets and so present a more challenging condition.

For the simulated data sets, we recorded the true species tree and true gene trees generated during the simulation process, which allows us to exactly quantify the topological error in the estimated tree (*34*). We computed the missing branch rate (also called the false-negative rate), which is the proportion of branches in the true tree that are missing from the estimated tree, as well as the false-positive rate, which is the proportion of branches in the estimated tree that

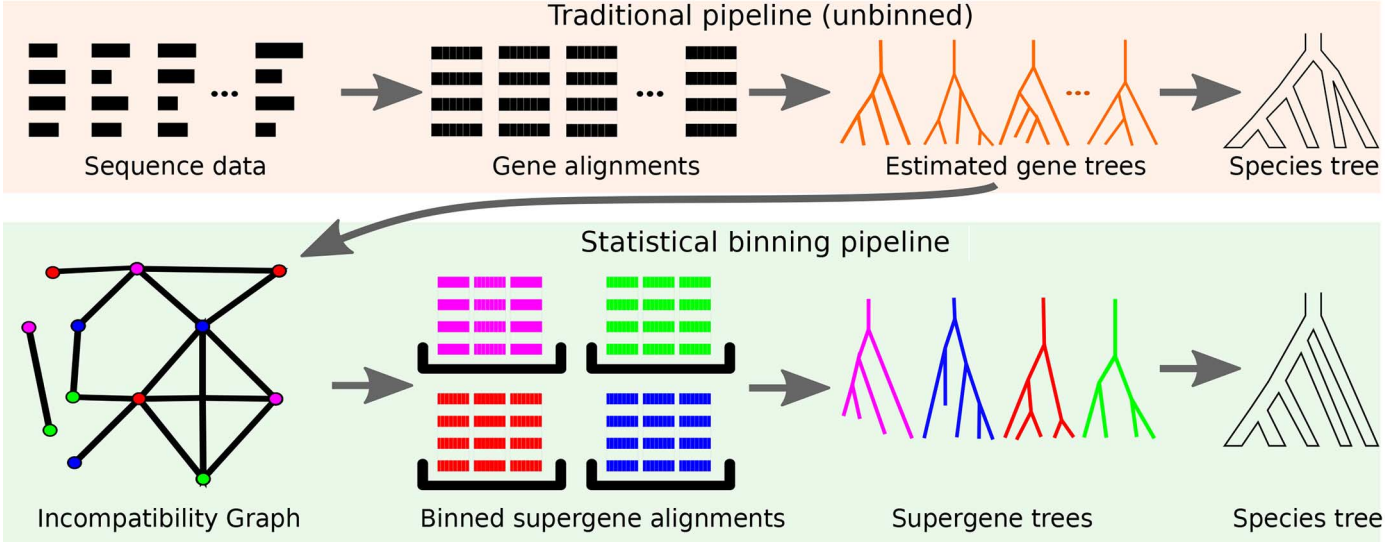


Fig. 1. Binning procedure. In traditional pipelines, gene trees are estimated from input sequence alignments and then combined into a species tree using a coalescent-based summary method. Statistical binning takes estimated gene trees and builds an incompatibility graph in which each node represents an estimated gene tree and each edge represents a detected incompatibility between two estimated gene trees at the specified statistical support threshold or higher. We use a heuristic we developed to color the nodes of the graph so that no two adjacent nodes have the same color and so that the color classes are of similar sizes. This coloring of the nodes defines a partition of genes into bins and ensures that no two genes with strongly supported conflict are put in the same bin. For each bin, individual gene alignments are concatenated to get a supergene alignment, from which a supergene tree is estimated using maximum-likelihood analysis. The supergene trees are then used as input to the summary method of choice to produce an estimated species tree.

are not present in the true tree. We measured how well the distribution of rooted gene trees is estimated by comparing triplet frequency distributions calculated from true gene trees and estimated gene trees; this is important because MP-EST uses estimated triplet distributions to construct the species tree. We measured estimation error in the species tree branch lengths as follows: Given a branch in an estimated species tree that is also present in the true species tree, we record the ratio of the length estimated for that branch by MP-EST to the true length of the branch (both in coalescent units) in the true (model) species tree (*34*).

Results

Unbinned MP-EST, binned MP-EST, and concatenation on 1000 avian genes with varying BS support in the gene trees are shown in Fig. 2A (see also table S2). Binned MP-EST was consistently and significantly more accurate than concatenation [$P < 10^{-5}$; all the statistical significance results reported henceforth are based on the two-way analysis of variance (ANOVA) test with Benjamini-Hochberg (BH) correction, and all the P values are reported in tables S4 and S5] and was also significantly more accurate than unbinned MP-EST ($P = 0.0001$). For gene trees with the highest BS values (i.e., long intron–like genes), both binned and unbinned MP-EST

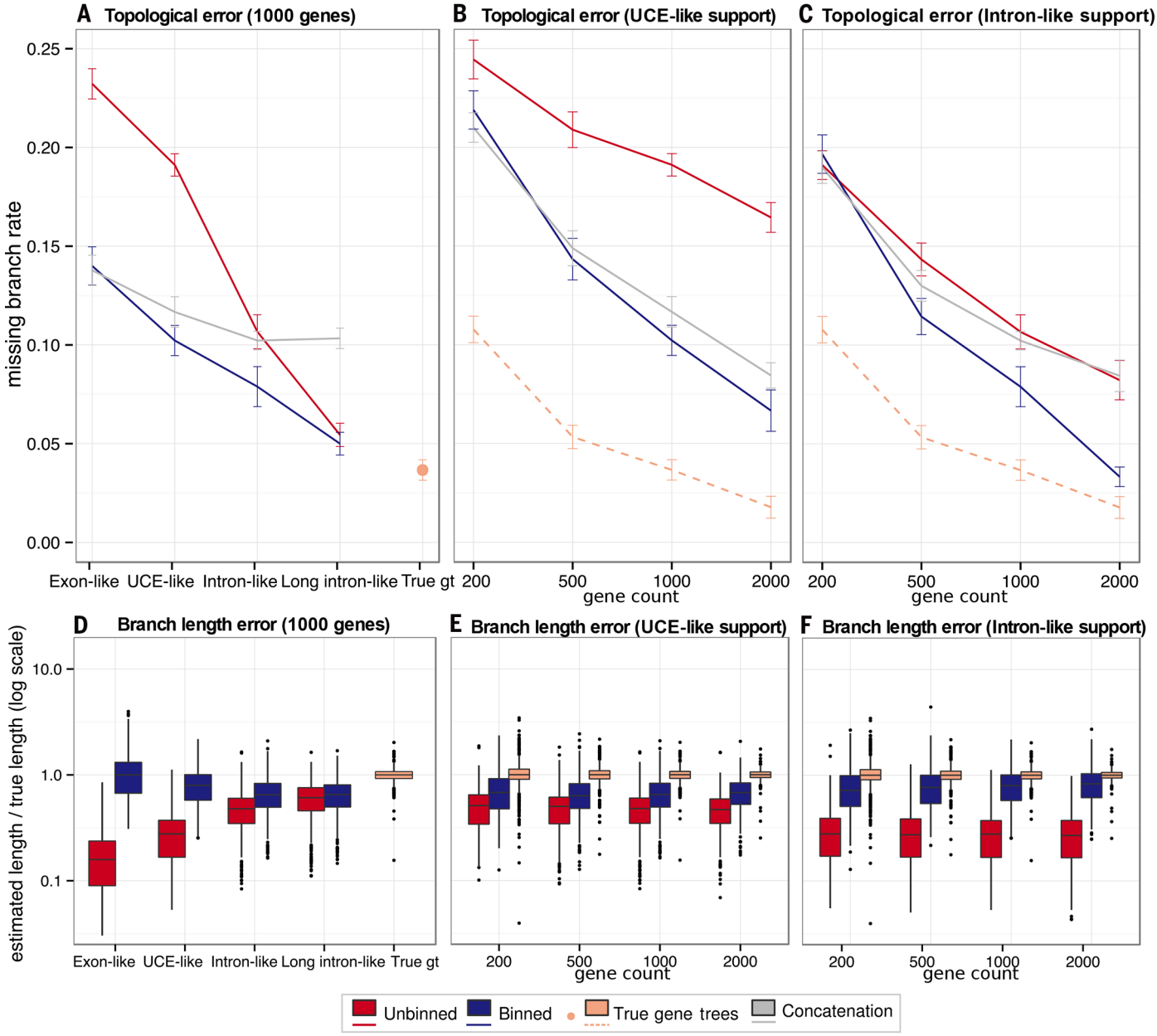


Fig. 2. Effect of binning on MP-EST on the simulated avian data sets with 1× ILS. (**A** to **C**) Species tree-topology error and (**D** to **F**) species tree branch–length error (boxplots with the ratio of estimated branch length to true branch length for branches of the true tree that appear in the estimated tree, 1 indicates correct estimation). MP-EST assigned an arbitrarily small length in the model tree to one branch, which we exclude from branch length calculations. (A) and (D) fix the number of genes to 1000 genes and vary gene tree support. The other panels fix gene tree support to UCE-like (B) and (E) and intronlike (C) and (F) and vary the number of genes. Results are over 10 replicates for the condition with 2000 genes, and 20 replicates for all other conditions. MP-EST has significantly lower topological error compared with MP-EST for (A), (B), and (C) ($P < 10^{-5}$, $P < 10^{-5}$, and $P = 0.002$, respectively), and concatenation for (A) ($P = 0.0001$) and (C) ($P = 0.01$). See also tables S2, S4, S5, and S6, and see fig. S7 for false-positive rates.

species trees had approximately the same error. However, as gene tree BS values decreased, the improvements obtained by binned MP-EST compared with unbinned MP-EST increased (P = 0.003 for the interaction effect; table S5). Concatenation was generally more accurate than unbinned MP-EST, except for gene trees with the highest BS. Results for MRP and Greedy showed similar trends (figs. S6 and S8; supplementary online text).

When loci exhibit BS corresponding to gene trees calculated on the UCE-like (moderate BS) or intronlike (high BS) markers (Fig. 2, B and C), binned MP-EST was more accurate than unbinned MP-EST ($P < 10^{-5}$ for UCE-like and P = 0.002 for intronlike markers). Furthermore, the advantage provided by binning increased with the number of UCE-like loci (the impact is significant with P = 0.003). Binned MP-EST tended to be more accurate than concatenation on both UCE-like and intronlike loci, but the differences are significant only for intronlike genes (P = 0.011). The improvement of binned MP-EST over concatenation appeared to increase with the number of intronlike loci, but the interaction effect is not significant (P = 0.087). Finally, on the mixed-model condition, concatenation and binned MP-EST each had 7% error, whereas all the other methods had at least 11% error (fig. S6). On the simulated mammalian data sets, binned MP-EST generally either matched or improved upon both unbinned MP-EST and concatenation (Fig. 3A). On the moderate (63%) BS trees, binned MP-EST and concatenation had close accuracy (with no significant differences), but unbinned MP-EST was significantly less accurate than binned MP-EST ($P < 10^{-5}$), and some conditions showed substantial differences (e.g., 800 loci). On higher BS (79%) loci, binned MP-EST was significantly more accurate than concatenation (P = 0.003), but there were no statistically significant differences between binned MP-EST and unbinned MP-EST. On the mixed-model condition, which most closely resembles the real mammalian data set in terms of the number of genes and gene tree support, binned MP-EST had only 1.8% error, concatenation had 3.7% error, and unbinned MP-EST had 4.6% error (Fig. 3C).

MP-EST always underestimated species tree branch lengths in coalescent units when analyzing estimated gene trees (in some cases by close to an order of magnitude), whereas the binned MP-EST trees had more accurate branch lengths (Figs. 2, D to F, and 3B). Because branch lengths are model parameters that determine the amount of ILS, underestimating branch lengths directly means overestimating ILS.

In the experiments where we varied the amount of ILS, binned MP-EST had lower average tree error than both unbinned MP-EST and concatenation, regardless of the amount of ILS (Fig. 4). The differences between binned and unbinned MP-EST are significant for both avian and mammalian data sets ($P < 10^{-5}$ and P = 0.0001, respectively), and differences between binned MP-EST and concatenation are significant on the avian data sets (P = 0.004). Furthermore, for

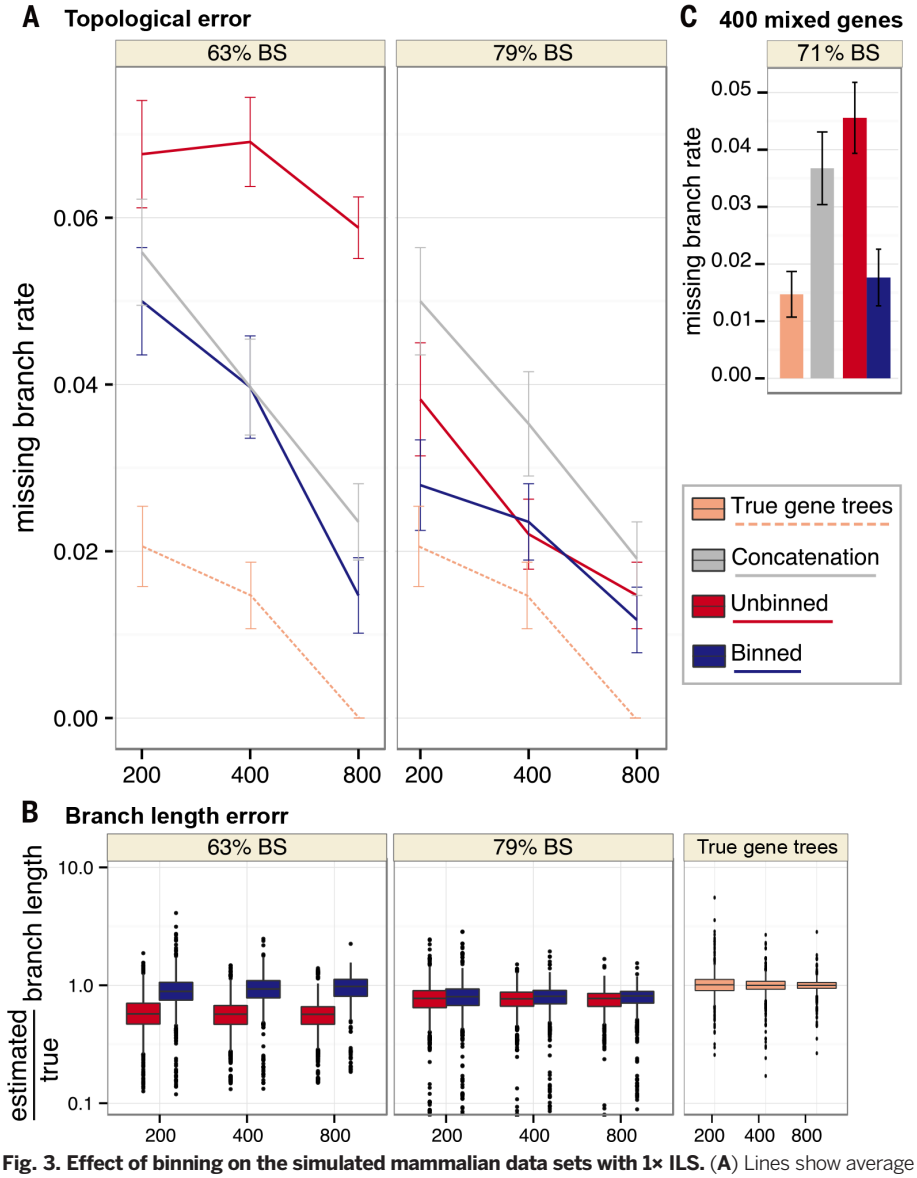


Fig. 3. Effect of binning on the simulated mammalian data sets with 1× ILS. (**A**) Lines show average topological tree error (missing branch rate) over 20 replicates for 200, 400, and 800 genes. Results are shown separately for gene trees with 63% and 79% BS. Binned MP-EST has significantly lower error compared to unbinned MP-EST for the 63% BS gene trees ($P < 10^{-5}$) and concatenation for the 79% BS gene trees (P = 0.003). (**B**) Error in branch lengths estimated by MP-EST in coalescent units is shown. (**C**) Topological error is shown for a mixed data set with 200 genes of 63% BS level and 200 genes of 79% BS level. See also table S3 and fig. S8.

the avian data sets, reducing the ILS level (2× condition) increased the impact of binning, and increasing the ILS level (0.5× condition) decreased the impact ($P < 10^{-5}$ for the interaction effect). The impact of ILS level on the mammalian data sets was similar but less pronounced and not statistically significant. For the reduced ILS (2×) models on both the avian and mammalian data sets, binned MP-EST was more accurate than unbinned MP-EST at estimating species tree topologies and branch lengths. For example, with 1000 UCE-like avian loci, unbinned MP-EST had 17.2% tree error, whereas binned MP-EST had only 5.9%. Performance on true gene trees provides an upper bound on what a summary method can achieve on estimated trees, and, as expected, MP-EST had its highest accuracy when run on true gene trees (Figs. 2 to 4).

Statistical binning improved the estimation of gene tree topologies (Table 1, but see also figs. S2 and S3), with the largest reductions in gene tree estimation error for the exonlike genes, and decreasing impact as the gene trees increased in BS. The reductions in triplet gene tree distribution estimation error were even larger (Table 1 and figs. S4 and S5), especially for loci with the lowest BS.

Biological data sets

We studied the avian data set (*31*) with 14,446 genes and 48 species, a mammalian data set with

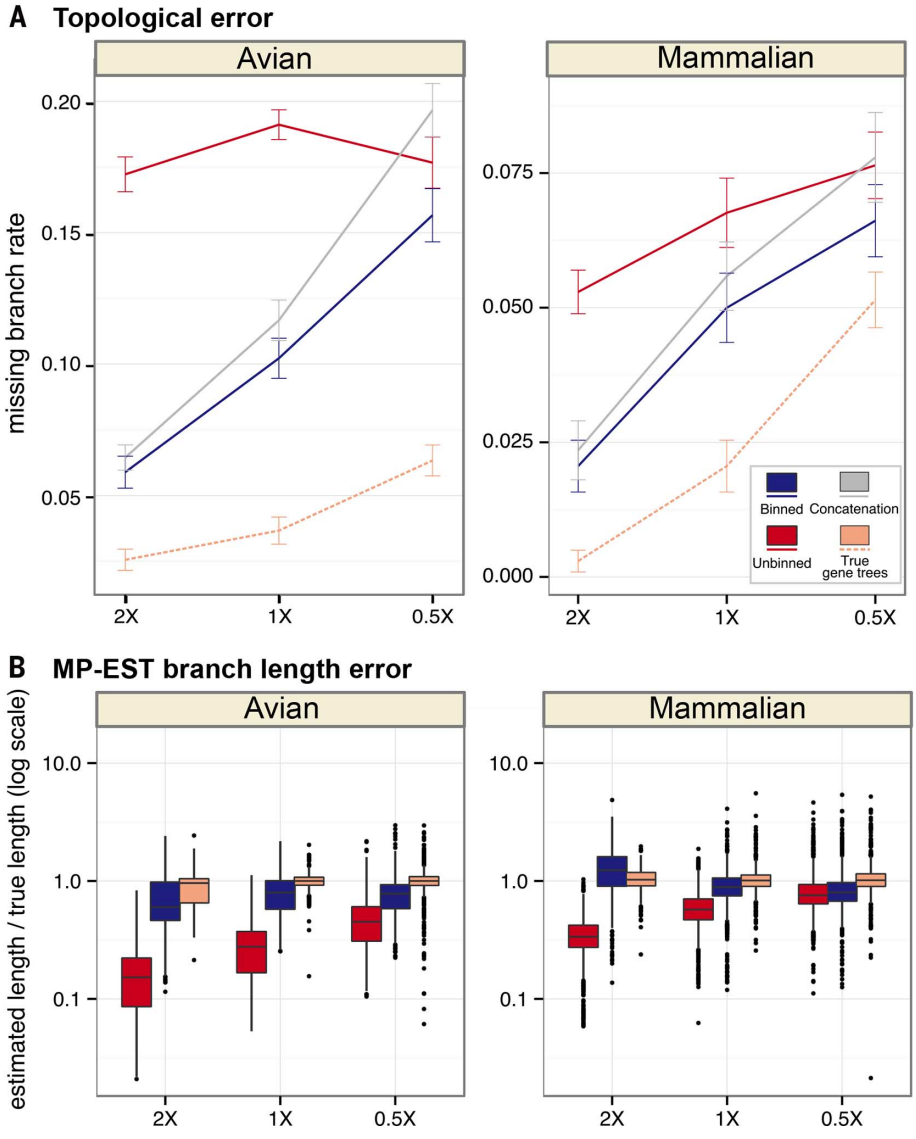


Fig. 4. Effects of ILS levels for the simulated avian and mammalian data sets. Levels of ILS are changed by multiplying all branch lengths in the model species tree by a factor of 0.5 (to increase ILS) or 2 (to reduce ILS). (**A**) Lines show average topological tree error (missing branch rate) over 20 replicates of 1000 UCE-like gene trees for avian and 200 gene trees with 63% BS for mammalian data sets. Binned MP-EST has significantly lower error compared with unbinned MP-EST for the avian ($P < 10^{-5}$) and the mammalian ($P = 0.0001$) data sets and concatenation for the avian data set ($P = 0.004$). (**B**) Error in branch lengths estimated by MP-EST in coalescent units.

447 genes and 37 species (*6*), yeast with 23 species and 1070 genes, vertebrates with 15 species and 1087 genes, and metazoa with 21 species and 225 genes (*32*). Each of these data sets shows evidence of gene tree discord (fig. S13), but they vary with respect to average BS (fig. S14). Species trees estimated using concatenation and gene trees were available (*31*, *32*), except for the maximum-likelihood gene trees from the mammalian data set, which we recomputed. We partitioned on the basis of individual loci when estimating supergene trees for the avian, metazoa, vertebrates, and yeast data sets (supplementary text). We present bin sizes (fig. S12), as well as a summary of the results obtained with binned and unbinned MP-EST on these data (Table 2).

Avian

The avian data set has very low average BS for almost all loci (fig. S1) and large topological distances between estimated gene trees. An unbinned MP-EST analysis of the full 14,446 loci produced a tree (Fig. 5A) with low to moderate support for some branches and failed to recover four key clades [Columbea (flamingo, grebe, pigeon, mesite, sandgrouse), Cursores (crane, killdeer), Otidimorphae (bustard, turaco, cuckoo), and Australaves (parrot, passerine, falcon, seriema)] that are recovered consistently in other analyses on the full genome data set (31), including concatenation analyses (fig. S17). Failure to recover Australaves is particularly surprising, as it has been recovered across different studies and types of data (*25*, *26*, *41*, *42*). In contrast, binned MP-EST on all 14,446 loci (Fig. 5A) had more highly supported branches and recovered all key clades.

An unbinned MP-EST tree generated on the introns-only data set (*31*) had 31 out of 45 edges with 100% support and 34 edges with 95% or higher BS; it also recovered all the key clades missing from the unbinned MP-EST tree computed on the full set of 14,446 loci. However, the binned MP-EST analysis (fig. S16) on the introns-only data set also recovered all the key clades and had higher support (33 edges with 100% support and 35 with 95% support or more), with increased support for some key novel clades (*31*).

Metazoa

The Metazoan data set also represents a challenging analysis, because the average BS is low (only 49%). The most important difference between the unbinned and binned MP-EST trees (Fig. 5B) is among Chordates, where the unbinned MP-EST tree put Cephalochordates (represented by *Branchiostoma floridae*) as sister to vertebrates (Craniates), and the binned MP-EST tree [as in the concatenation analysis, (fig. S20)], put Urochordates (represented by *Ciona intestinalis*) as sister to vertebrates. Although Cephalochordates were traditionally thought to be the sister to all the extant vertebrates (*43*), recent evidence supports Urochordates as the sister to all vertebrates (*44*–*46*), and hence, the binned MP-EST tree is likely correct. There are also some differences between the two trees within Protostomia, but both MP-EST trees had low support for those relations, and neither was congruent with the literature (supplementary text).

Mammalian

The mammalian data set has gene trees with substantially higher average BS (71%) but also demonstrates substantial gene tree incongruence. Differences between MP-EST and concatenation (using maximum likelihood) were observed for tree shrews and bats: The concatenated analysis put Scandentia (tree shrews) as sister to Glires (Rodentia/Lagomorpha), whereas the MP-EST analysis put Scandentia as sister to primates (*6*). We reanalyzed this data set and identified 21 loci with mislabeled sequences (subsequently confirmed by the authors) plus two outlier loci (fig. S18 and supplementary text). We removed these 23 loci and reanalyzed the data using concatenation and both binned and unbinned MP-EST (fig. S19). We recovered a concatenation tree topologically identical to the concatenation tree in (*6*). The unbinned MP-EST tree on this reduced gene set was similar to the unbinned MP-EST tree reported in (*6*) but had lower support for tree shrews as sister to primates [99% in (*6*), 64% with our analysis], and there was

Table 1. Gene tree estimation error, with and without binning for simulated data sets. Results are shown for fixed number of genes (1000 for avian and 200 for mammalian) and levels of ILS (1×, i.e., observed), but also see figs. S2 to S5. Individual gene tree (GT) error is mean topological distance, measured using the missing branch rate between the true gene tree and all 200 bootstrap replicates of each estimated gene tree. For the supergene trees, each bootstrap replicate of each supergene tree is compared separately against each true gene tree for the genes put in that bin. We also characterize gene tree distributions by calculating the triplet frequencies for all possible triplets, and we do this both for true and estimated gene trees (using all 200 bootstrap replicates of all genes and supergenes in the case of estimated trees). Thus, we obtain a true and an estimated triplet frequency distribution for each of the triplets. We report the mean Kullback-Leibler (KL) divergence of the estimated distribution from the true distribution.

Tree	Genomic markers	Length (bp)	Individual GT error (%)		GT distribution error (KL)	
			Unbinned	Binned	Unbinned	Binned
Avian (1000 genes)	Exonlike	250	79	57	0.234	0.025
	UCE-like	500	69	57	0.120	0.008
	Intronlike	1000	55	51	0.033	0.008
	Long intron–like	1500	46	45	0.011	0.007
Mammalian (200 genes)	63% BS	500	43	35	0.119	0.019
	79% BS	1000	27	26	0.038	0.027

Table 2. Results on the biological data sets. We compare MP-EST trees in terms of BS (the number of edges with BS equal to 100%, edges with BS at least 95%, and average BS), the distance between concatenation (DC) and MP-EST trees (number of missing branches), and with respect to biologically interesting differences between the two trees. The total number of branches in each tree is given parenthetically in the first column.

Tree (branches)	Gene trees	Bootstrap Support			DC	Interesting clades
		100%	>95%	Mean		
Avian (45)	Unbinned	29	34	0.95	12	Did not recover Australaves, Columbea, Cursores, and Otidimorphae, all recovered by other phylogenomic analysis
	Binned-50%	36	39	0.96	5	
Mammals (34)	Unbinned	30	30	0.98	2	Recovers Scandentia/Primates
	Binned-75%	30	31	0.98	1	Recovers Scandentia/Glires
Metazoa (18)	Unbinned	10	10	0.83	5	Rejects Olfactores (urochordates/vertebrates) and Eumetazoa
	Binned-75%	10	12	0.89	2	Rejects Eumetazoa
Vertebrates (15)	Unbinned	14	15	1	0	
	Binned-50%	14	14	0.99	0	
Yeast (20)	Unbinned	19	20	1	1	
	Binned-50%	19	19	0.98	1	

one topological difference among low support edges; the exact cause of these differences is not clear to us.

The binned MP-EST and unbinned MP-EST trees on the reduced gene set were very similar, but tree shrews were sister to Glires with 80% support in the binned MP-EST tree, just as their position was recovered in the concatenation tree. Thus, the placement of Scandentia, and whether it is sister to primates or to Glires, depends on the mode of analysis. This agreement between the binned MP-EST analysis and concatenated analysis of the reduced data set may be an important finding, but contradicts Janecka *et al.* (*47*) (which specifically addressed this question) and (*48*). However, these two studies did not use coalescent-based methods to estimate species trees. The unbinned and binned MP-EST trees placed bats identically as sister to all other Laurasitheria (except for the basal Eulipotyphyla) and so differed from the concatenation tree with respect to bats.

Vertebrates

The vertebrate data set had the highest average BS (76%) of all data sets we examined. Binned and unbinned MP-EST trees had the same topology, and both were topologically identical to the concatenation tree on the same data (fig. S21 and supplementary text).

Yeast

The yeast data set has relatively high average BS (72%). The binned and unbinned MP-EST topologies were identical, and both had 100% support for all but one branch (fig. S22).

Discussion

Our simulation results demonstrate that binning reduces error in estimated species tree topologies and branch lengths, gene tree topologies, and gene tree distributions under the conditions we studied. These reductions in error result in estimations of ILS that are closer to correct ILS levels than unbinned MP-EST, which tends to overestimate ILS levels. In our analyses, although unbinned methods are rarely more accurate than concatenation, binned MP-EST is almost always at least as accurate as concatenation, and there are many model conditions in which binned MP-EST is more accurate than concatenation, whereas unbinned MP-EST is less accurate than concatenation.

The biological data sets examined here show that binning affects analyses of data sets with less well supported gene trees (avian and metazoa) and has little impact on the yeast and vertebrate data sets, both of which have very well resolved gene trees. Notably, binning impacts the MP-EST analysis of the mammalian data set, which also has fairly well resolved gene trees. Where binning has an impact, binned MP-EST typically produces trees in closer agreement with accepted reconstructions than unbinned MP-EST. Binning reduces gene tree incongruence on biological data sets (fig. S13), which suggests that binned MP-EST may not overestimate ILS on biological data sets as much as unbinned MP-EST does. This trend is consistent with performance on simulated data and suggests that more-accurate estimates of ILS in biological data may be obtained through the use of statistical binning.

As binning can group genes together with different true topologies, it can result in misspecified models in the supergene tree estimation step and can reduce the accuracy of the estimated gene tree distributions. However, our simulations suggest that estimated gene tree distributions are more accurate after binning.

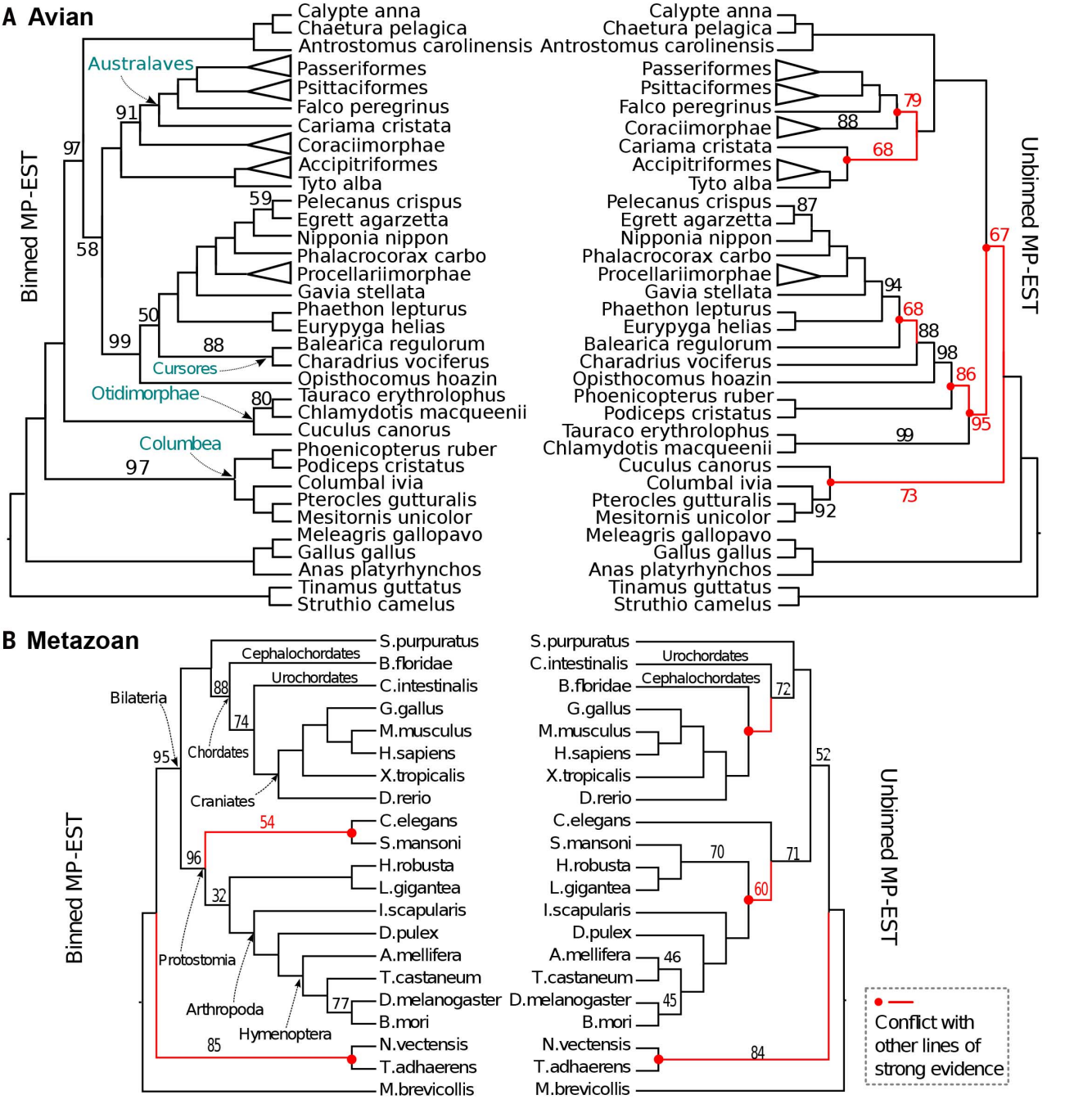


Fig. 5. Results on the (A) avian and (B) metazoan biological data sets using binned and unbinned MP-EST. Branches without designation represent 100% support.

We suggest that this is because binning will never group genes with different topologies together unless the conflicting branches had low support, which likely results from insufficient phylogenetic signal. As we have shown, the inclusion of poorly estimated gene trees distorts the estimated triplet gene tree distribution, and binning reduces this noise, which suggests that the overall impact of binning is beneficial. These results are also consistent with the observation that coalescent-based summary methods can be robust to recombination (*49*).

Our study explored gene tree estimation error arising from insufficient phylogenetic signal in the gene sequences; however, gene tree estimation error can also come from poorly estimated alignments (*50*) or errors introduced during the tree inference (*51*, *52*). Because our studies focused on insufficient phylogenetic signal, we have no evidence that binning could reduce phylogenetic error due to alignment error or misspecification for the sequence evolution model. Consequently, appropriate care should be devoted to obtaining good alignments and choosing an adequate model of sequence evolution to reconstruct both gene and supergene trees.

In our simulation, we only allowed ILS as a source of discord between true gene trees and true species trees; hence, these model conditions favor MP-EST (which is based on the same model used for simulations) over concatenation (which assumes no ILS is present). Given this, the fact that unbinned MP-EST is less accurate than concatenation in many conditions is noteworthy. Future studies based on model conditions in which other sources of gene tree discord

(e.g., duplication and loss, incorrect orthology assessments, recombination, introgression, horizontal gene transfer, and hybridization) are included would enable a better understanding of the relative accuracy of concatenation and coalescent-based species tree estimation and the impact of using binning under those conditions.

Statistical binning is just one step in a pipeline that begins with sequence alignments on different loci and ends with an estimated species tree, and variations of this pipeline might lead to improved accuracy. For example, gene trees could be estimated using Bayesian methods instead of maximum likelihood (*53*), and rigorous statistical tests for combinability (*54*) could be used instead of bootstrap branch support values; both variants might improve species tree estimations but would result in substantially increased running time. Bins could also be created without attempting to produce balanced sizes; in which case, bins could be weighted by their size. However, binning without attempting to evaluate whether genes have a common tree [as used in the "naïve binning" technique (*28*)] may not provide the improvements in accuracy seen here. Statistical binning is not likely to be useful for methods that estimate species trees directly from sequence data [e.g., (*21*–*23*)], because binning reduces the amount of data given to the method and can only be beneficial if it also improves the quality of the input data.

Our study demonstrates that binning is useful when the input is a set of estimated gene trees, because the method typically reduces the estimation error in the gene trees; however, binning sites together cannot improve the quality of sequence data. Furthermore, binning cannot be used with methods (like SNAPP) that are restricted to biallelic markers, because binning biallelic markers would create markers exhibiting at least three states and thus would violate that property.

The algorithmic techniques used in statistical binning are fast enough to use on data sets with many thousands of loci (such as the avian phylogenomics data set). Thus, statistical binning enables coalescent-based methods to be used on genome-scale data and can help to resolve challenging phylogenetic questions, including the avian Tree of Life.

Methods

The statistical binning technique includes a combinability test that evaluates whether a given pair of genes is likely to have significant topological incongruence, so that a concatenated analysis of those two genes is likely to be problematic. Because supergene trees can be estimated using partitioned concatenated analyses (which would allow the branch lengths and other model parameters to be reestimated for each gene within a partition), we only need to consider topological incongruence. Thus, we can place genes with the same true gene tree topology in the same bin, even if the trees differ in other respects (e.g., branch lengths).

We use maximum likelihood with bootstrapping to estimate gene trees with branch support values, and we say that a given pair of trees exhibits conflict at threshold t if there is a pair of incompatible branches (meaning they cannot coexist in any tree), one in each of the two gene trees, each with BS of at least t. Two trees that do not exhibit conflict at threshold t are combinable, and a set of trees for which all pairs are combinable is a combinable set.

Saying that two branches are incompatible means that no tree can be constructed that has both of these branches (*55*) (more specifically, no tree exists with branches that induce the bipartitions defined by these two branches). Thus, to test two trees for incompatibility at threshold t or higher, we collapse all branches in each tree with support below t, and then ask whether a tree exists that is a common refinement of these two collapsed trees. Testing for compatibility of two trees can be performed in linear time (*55*); hence, this calculation is fast.

The partitioning step uses a graph-based optimization, in which we build a graph in which each gene is represented by a node and an edge is present between two nodes (i.e., genes) if the estimated trees on that pair of genes exhibit conflict at threshold t. By definition, the graph depends on the parameter t; thus, smaller values for t will generally consider trees less likely to be combinable than larger values.

The graph created is called an incompatibility graph. To create bins from this graph, we color the nodes of the graph so that no two nodes with the same color are adjacent, and put all nodes with the same color into a common bin. Each bin thus contains a set of genes where no pairwise incompatibility has support of t or greater; hence, adjusting the support threshold t enables more aggressive or conservative binning. Once bins are formed, alignments of genes in the same bin are concatenated into a supergene alignment, and supergene trees are estimated on these supergene alignments using maximum likelihood. These supergene trees are used as input to the summary method of choice.

Because the statistically consistent methods use the distribution of gene trees to estimate the species tree, it is important for the supergene tree distribution to be close to the gene tree distribution; for this reason, we seek a node coloring in which the different color classes have approximately the same size (i.e., are balanced). We also seek a node coloring with a small number of colors, so that we have the largest bins we can, given the constraints imposed by combinability. However, finding a minimum node coloring (regardless of whether bins are balanced) is NP-hard and, consequently, is believed to not be solvable in polynomial time (*56*). Therefore, we developed a heuristic algorithm [based on the Brélaz heuristic (*56*) (see fig. S10)] for finding the smallest number of balanced bins (*34*). Our greedy heuristic algorithm processes genes one by one in a particular order and adds each gene to the smallest bin that has no incompatibility with it (*34*). When two or more bins have the same smallest size, the algorithm breaks the ties arbitrarily. Our experiments indicate that variations causes by these arbitrary choices mostly affect low support branches (see fig. S11), but also that a consensus tree from multiple runs of the binning approach can have higher average accuracy than individual runs (fig. S11). Therefore, if computational resources permit, we recommend that several runs of statistical binning be applied (each breaking ties differently) to produce several different estimated species trees. These trees can be compared with each other, to explore sensitivity and reliability, or a consensus tree of these trees can be used as a point estimate of the species tree.

We set the statistical support threshold t as follows. We note that using 75% for the BS has been a standard threshold for branch reliability (*57*), and so 75% represents a reasonable setting for t; however, when the data sets are large, we can afford to be more conservative and pick a smaller threshold. We also explored the effect of the support threshold (fig. S9) and saw that setting t to either 50 or 75% gave good results. Therefore, we set two thresholds: a conservative threshold of t = 50% that we use for data sets with at least 1000 genes and a moderate threshold of t = 75% that we use for the other data sets.

REFERENCES AND NOTES

1. A. Rokas, B. L. Williams, N. King, S. B. Carroll, Genome-scale approaches to resolving incongruence in molecular phylogenies. *Nature* **425**, 798–804 (2003). doi: 10.1038/nature02053; pmid: 14574403
2. W. P. Maddison, Gene trees in species trees. *Syst. Biol.* **46**, 523–536 (1997). doi: 10.1093/sysbio/46.3.523
3. S. V. Edwards, Is a new and general theory of molecular systematics emerging? *Evolution* **63**, 1–19 (2009). doi: 10.1111/j.1558-5646.2008.00549.x; pmid: 19146594
4. L. L. Knowles, Estimating species trees: Methods of phylogenetic analysis when there is incongruence across genes. *Syst. Biol.* **58**, 463–467 (2009). doi: 10.1093/sysbio/syp061; pmid: 20525600
5. S. J. Hackett *et al.*, A phylogenomic study of birds reveals their evolutionary history. *Science* **320**, 1763–1768 (2008). doi: 10.1126/science.1157704; pmid: 18583609
6. S. Song, L. Liu, S. V. Edwards, S. Wu, Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 14942–14947 (2012). doi: 10.1073/pnas.1211733109; pmid: 22930817
7. R. W. Meredith *et al.*, Impacts of the Cretaceous Terrestrial Revolution and KPg extinction on mammal diversification. *Science* **334**, 521–524 (2011). doi: 10.1126/science.1211028; pmid: 21940861
8. J. H. Degnan, N. A. Rosenberg, Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol. Evol.* **24**, 332–340 (2009). doi: 10.1016/j.tree.2009.01.009; pmid: 19307040
9. N. A. Rosenberg, Discordance of species trees with their most likely gene trees: A unifying principle. *Mol. Biol. Evol.* **30**, 2709–2713 (2013). doi: 10.1093/molbev/mst160; pmid: 24030555
10. S. V. Edwards, L. Liu, D. K. Pearl, High-resolution species trees without concatenation. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 5936–5941 (2007). doi: 10.1073/pnas.0607004104; pmid: 17392434
11. L. S. Kubatko, J. H. Degnan, Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Syst. Biol.* **56**, 17–24 (2007). doi: 10.1080/10635150601146041; pmid: 17366134
12. B. R. Larget, S. K. Kotha, C. N. Dewey, C. Ané, BUCKy: Gene tree/species tree reconciliation with Bayesian concordance analysis. *Bioinformatics* **26**, 2910–2911 (2010). doi: 10.1093/bioinformatics/btq539; pmid: 20861028

13. L. Liu, L. Yu, S. V. Edwards, A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. *BMC Evol. Biol.* **10**, 302 (2010). doi: 10.1186/1471-2148-10-302; pmid: 20937096
14. L. Liu, D. K. Pearl, Species trees from gene trees: Reconstructing Bayesian posterior distributions of a species phylogeny using estimated gene tree distributions. *Syst. Biol.* **56**, 504–514 (2007). doi: 10.1080/10635150701429982; pmid: 17562474
15. L. Liu, BEST: Bayesian estimation of species trees under the coalescent model. *Bioinformatics* **24**, 2542–2543 (2008). doi: 10.1093/bioinformatics/btn484; pmid: 18799483
16. L. Liu, L. Yu, D. K. Pearl, S. V. Edwards, Estimating species phylogenies using coalescence times among sequences. *Syst. Biol.* **58**, 468–477 (2009). doi: 10.1093/sysbio/syp031; pmid: 20525601
17. E. Mossel, S. Roch, Incomplete lineage sorting: Consistent phylogeny estimation from multiple loci. *IEEE/ACM Trans. Comp. Biol. Bioinform.* **7**, 166–171 (2010). doi: 10.1109/TCBB.2008.66; pmid: 20150678
18. J. Heled, A. J. Drummond, Bayesian inference of species trees from multilocus data. *Mol. Biol. Evol.* **27**, 570–580 (2010). doi: 10.1093/molbev/msp274; pmid: 19906793
19. Y. Yu, T. Warnow, L. Nakhleh, Algorithms for MDC-based multi-locus phylogeny inference: Beyond rooted binary gene trees on single alleles. *J. Comput. Biol.* **18**, 1543–1559 (2011). doi: 10.1089/cmb.2011.0174; pmid: 22035329
20. B. Rannala, Z. Yang, Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. *Genetics* **164**, 1645–1656 (2003). pmid: 12930768
21. D. Bryant, R. Bouckaert, J. Felsenstein, N. A. Rosenberg, A. RoyChoudhury, Inferring species trees directly from biallelic genetic markers: Bypassing gene trees in a full coalescent analysis. *Mol. Biol. Evol.* **29**, 1917–1932 (2012). doi: 10.1093/molbev/mss086; pmid: 22422763
22. N. De Maio, C. Schlötterer, C. Kosiol, Linking great apes genome evolution across time scales using polymorphism-aware phylogenetic models. *Mol. Biol. Evol.* **30**, 2249–2262 (2013). doi: 10.1093/molbev/mst131; pmid: 23906727
23. J. Chifman, L. Kubatko, Quartet inference from SNP data under the coalescent model. *Bioinformatics* (2014). doi: 10.1093/bioinformatics/btu530
24. B. T. Smith, M. G. Harvey, B. C. Faircloth, T. C. Glenn, R. T. Brumfield, Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. *Syst. Biol.* **63**, 83–95 (2014). doi: 10.1093/sysbio/syt061; pmid: 24021724
25. R. T. Kimball, N. Wang, V. Heimer-McGinn, C. Ferguson, E. L. Braun, Identifying localized biases in large datasets: A case study using the avian tree of life. *Mol. Phylogenet. Evol.* **69**, 1021–1032 (2013). doi: 10.1016/j.ympev.2013.05.029; pmid: 23791948
26. J. E. McCormack *et al.*, A phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput sequencing. *PLOS ONE* **8**, e54848 (2013). doi: 10.1371/journal.pone.0054848; pmid: 23382987
27. M. DeGiorgio, J. H. Degnan, Fast and consistent estimation of species trees using supermatrix rooted triples. *Mol. Biol. Evol.* **27**, 552–569 (2010). doi: 10.1093/molbev/msp250; pmid: 19833741
28. M. S. Bayzid, T. Warnow, Naive binning improves phylogenomic analyses. *Bioinformatics* **29**, 2277–2284 (2013). doi: 10.1093/bioinformatics/btt394; pmid: 23842808
29. S. Patel, R. Kimball, E. Braun, Error in phylogenetic estimation for bushes in the Tree of Life. *J. Phylogenet. Evol. Biol.* **1**, 2 (2013). doi: 10.4172/2329-9002.1000110
30. L. Nakhleh, U. Roshan, K. St. John, J. Sun, T. Warnow, Designing fast converging phylogenetic methods. *Bioinformatics* **17** (suppl. 1), S190–S198 (2001). doi: 10.1093/bioinformatics/17.suppl_1.S190; pmid: 11473009
31. E. D. Jarvis *et al.*, Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* **346**, 1320–1331 (2014).
32. L. Salichos, A. Rokas, Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature* **497**, 327–331 (2013). doi: 10.1038/nature12130; pmid: 23657258
33. A. Stamatakis, RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006). doi: 10.1093/bioinformatics/btl446; pmid: 16928733
34. Materials and methods are available as supplementary material on *Science* Online.
35. M. A. Ragan, Phylogenetic inference based on matrix representation of trees. *Mol. Phylogenet. Evol.* **1**, 53–58 (1992). doi: 10.1016/1055-7903(92)90035-F; pmid: 1342924
36. T.-K. Seo, Calculating bootstrap probabilities of phylogeny using multilocus sequence data. *Mol. Biol. Evol.* **25**, 960–971 (2008). doi: 10.1093/molbev/msn043; pmid: 18281270
37. A. D. Leaché, R. B. Harris, B. Rannala, Z. Yang, The influence of gene flow on species tree estimation: A simulation study. *Syst. Biol.* **63**, 17–30 (2013). doi: 10.1093/sysbio/syt049; pmid: 23939193
38. L. Zhao *et al.*, Phylogenomic analyses of nuclear genes reveal the evolutionary relationships within the BEP clade and the evidence of positive selection in Poaceae. *PLOS ONE* **8**, e64642 (2013). doi: 10.1371/journal.pone.0064642; pmid: 23734211
39. B. Zhong, L. Liu, Z. Yan, D. Penny, Origin of land plants using the multispecies coalescent model. *Trends Plant Sci.* **18**, 492–495 (2013). doi: 10.1016/j.tplants.2013.04.009;pmid: 23707196
40. J. H. Degnan, M. DeGiorgio, D. Bryant, N. A. Rosenberg, Properties of consensus methods for inferring species trees from gene trees. *Syst. Biol.* **58**, 35–54 (2009). doi: 10.1093/sysbio/syp008; pmid: 20525567
41. A. Suh *et al.*, Mesozoic retroposons reveal parrots as the closest living relatives of passerine birds. *Nat. Commun.* **2**, 443 (2011). doi: 10.1038/ncomms1448; pmid: 21863010
42. N. Wang, E. L. Braun, R. T. Kimball, Testing hypotheses about the sister group of the passeriformes using an independent 30-locus data set. *Mol. Biol. Evol.* **29**, 737–750 (2012). doi: 10.1093/molbev/msr230; pmid: 21940640
43. C. Nielsen, *Animal Evolution: Interrelationships of the Living Phyla* (Oxford Univer Press, Oxford, 2012).
44. F. Delsuc, H. Brinkmann, D. Chourrout, H. Philippe, Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* **439**, 965–968 (2006). doi: 10.1038/nature04336; pmid: 16495997
45. S. J. Bourlat *et al.*, Deuterostome phylogeny reveals monophyletic chordates and the new phylum Xenoturbellida. *Nature* **444**, 85–88 (2006). doi: 10.1038/nature05241; pmid: 17051155
46. T. R. Singh *et al.*, Tunicate mitogenomics and phylogenetics: Peculiarities of the *Herdmania momus* mitochondrial genome and support for the new chordate phylogeny. *BMC Genomics* **10**, 534 (2009). doi: 10.1186/1471-2164-10-534; pmid: 19922605
47. J. E. Janecka *et al.*, Molecular and genomic data identify the closest living relative of primates. *Science* **318**, 792–794 (2007). doi: 10.1126/science.1147555; pmid: 17975064
48. B. Boussau *et al.*, Genome-scale coestimation of species and gene trees. *Genome Res.* **23**, 323–330 (2013). doi: 10.1101/gr.141978.112; pmid: 23132911
49. H. C. Lanier, L. L. Knowles, Is recombination a problem for species-tree analyses? *Syst. Biol.* **61**, 691–701 (2012). doi: 10.1093/sysbio/syr128; pmid: 22215721
50. K. Liu, S. Raghavan, S. Nelesen, C. R. Linder, T. Warnow, Rapid and accurate large-scale coestimation of sequence alignments and phylogenetic trees. *Science* **324**, 1561–1564 (2009). doi: 10.1126/science.1171243; pmid: 19541996
51. J. Felsenstein, Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* **27**, 401 (1978). doi: 10.2307/2412923
52. W. G. Weisburg, S. J. Giovannoni, C. R. Woese, The Deinococcus-Thermus phylum and the effect of rRNA composition on phylogenetic tree construction. *Syst. Appl. Microbiol.* **11**, 128–134 (1989). doi: 10.1016/S0723-2020(89)80051-7; pmid: 11542160
53. M. DeGiorgio, J. H. Degnan, Robustness to divergence time underestimation when inferring species trees from estimated gene trees. *Syst. Biol.* **63**, 66–82 (2014). doi: 10.1093/sysbio/syt059; pmid: 23988674
54. J. W. Leigh, E. Susko, M. Baumgartner, A. J. Roger, Testing congruence in phylogenomic analysis. *Syst. Biol.* **57**, 104–115 (2008).doi: 10.1080/10635150801910436; pmid: 18288620
55. T. Warnow, Tree compatibility and inferring evolutionary history. *J. Algorithms* **16**, 388–407 (1994). doi: 10.1006/jagm.1994.1018
56. D. Brélaz, New methods to color the vertices of a graph. *Commun. ACM* **22**, 251–256 (1979). doi: 10.1145/359094.359101
57. T. P. Wilcox, D. J. Zwickl, T. A. Heath, D. M. Hillis, Phylogenetic relationships of the dwarf boas and a comparison of Bayesian and bootstrap measures of phylogenetic support. *Mol. Phylogenet. Evol.* **25**, 361–371 (2002). doi: 10.1016/S1055-7903(02)00244-0; pmid: 12414316

ACKNOWLEDGMENTS

SM was supported by a graduate fellowship from Howard Hughes Medical Institute (HHMI) M.S.B. was supported by a graduate fellowship from the Fulbright Foundation, by NSF under grant DBI-10735191 to iPLANT, and by the University of Alberta, made possible through a donation from Musea Ventures, which is held by G. K.-S. Wong; T.W. was supported by NSF under grant DBI-1062335; and B.B. was supported by a postdoctoral fellowship from the Human Frontier Science Program, the CNRS, and Agence Nationale de la Recherche (ANR) through grant ANR-10-BINF-01-01 Ancestrome. The authors thank E. Braun, G. Chacko, S. Edwards, T. Gilbert, E. Jarvis, L. Liu, D. Posada, and G. Zhang for helpful comments. The computational analyses in this study were performed using generous allocations on TACC (both Lonestar and Stampede) and on the Condor cluster of the University of Texas Computer Science department. All the data sets used in the paper and the software for performing statistical binning can be found at http://www.ideals.illinois.edu/handle/2142/55319.

SUPPLEMENTARY MATERIALS

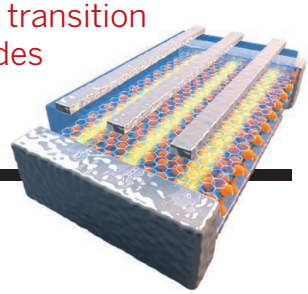
www.sciencemag.org/content/346/6215/1250463/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S22
Tables S1 to S6
References (*58*–*70*)

6 January 2014; accepted 3 September 2014
10.1126/science.1250463

RESEARCH

Exotic properties of transition metal dichalcogenides

Qian et al., *p. 1344*



IN *SCIENCE* JOURNALS

Edited by **Stella Hurtley**



Bees pollinate lavender and other flowering plants

POLLINATOR DECLINES

"Green" revolution not so great for pollinators

It is increasingly recognized that many pollinator populations are declining. Ollerton *et al.* looked at British historical distribution records for bees and flower-visiting wasps across the past century. Though it is well known that agricultural intensification after World War II had a negative impact on many species, pollinator declines began in the decades preceding this time, when other changes were made to agricultural practices and policies. — SNV

Science, this issue p. 1360

ULTRAFAST DYNAMICS

Watching electrons dart through silicon

The ultimate speed limit in electronic circuitry is set by the motion of the electrons themselves. Schultze *et al.* applied attosecond spectroscopy to glimpse this motion in a sample of silicon, the semiconducting building block of modern integrated circuits (see the Perspective by Spielmann). The technique distinguished the electron dynamics—which proceed faster than a quadrillionth of a second after laser excitation—from the comparatively slower lattice motion of the silicon atomic nuclei. — JSY

Science, this issue p. 1348; see also p. 1293

SEPARATION MEMBRANES

Metal-organic framework material membranes

There continues to be a lot of interest in developing membranes for gas separations that go beyond the current polymer membranes used commercially for this purpose. Peng *et al.* took a porous metal-organic framework material with a layered structure and exfoliated it to give nanometer-thick molecular sieves. The membranes were exceptionally good at separating hydrogen gas from carbon dioxide both in terms of permeance and selectivity. — MSL

Science, this issue p. 1356

HIV ANTIBODIES

A fierce competition to attack HIV

The Holy Grail in the HIV field is a protective vaccine. However, because HIV mutates rapidly, it can escape from vaccine-elicited antibodies. Nevertheless, some HIV-positive individuals harbor antibodies that HIV cannot escape from very easily, so-called broadly neutralizing antibodies (bNAbs). McGuire *et al.* asked why these bNAbs don't win out over their more narrow brethren (nNAbs) early after infection (see the Perspective by Ofek and Diskin). They compared the responses of B cells carrying germline versions of nNAbs and bNAbs—that is, versions of these Abs that HIV would initially encounter—to HIV's envelope glycoprotein (Env). The nNAbs could recognize Env from a more diverse array of viruses compared to bNAbs, which puts them at a competitive advantage. — KLM

Science, this issue p. 1380; see also p. 1290

NUTRITIONAL IMMUNITY

The frontline of host-pathogen coevolution

Pathogens have to subvert a host's innate defenses to avoid being killed. Barber and Elde now show that this principle extends

Downloaded from www.sciencemag.org on December 14, 2014

CREDITS: (TOP TO BOTTOM) YAN LIANG AND DEMIN LIU; SAGE_SOLAR/FLICKR

Published by AAAS

to nutrient-transporting proteins, such as transferrin, which binds iron (see the Perspective by Armitage and Drakesmith). Without iron, invading pathogens cannot replicate, but iron is sequestered in transferrin, which stops pathogens using it. So pathogens have evolved a succession of transporters that can hijack transferrin's iron. Over time, the primate transferrin binding surface has coevolved to wrestle iron back from the grip of pathogens. — CA

Science, this issue p. 1362; see also p. 1299

POLITICAL SCIENCE

Dialogue opens the door to attitude change

Personal contact between in-group and out-group individuals of equivalent status can reduce perceived differences and thus improve intergroup relations. LaCour and Green demonstrate that simply a 20-minute conversation with a gay canvasser produced a large and sustained shift in attitudes toward same-sex marriage for Los Angeles County residents. Surveys showed persistent change up to 9 months after the initial conversation. Indeed, the magnitude of the shift for the person who answered the door was as large as the difference between attitudes in Georgia and Massachusetts. — GJC

Science, this issue p. 1366

NANOLITHOGRAPHY

Laser shock imprinting for patterning metals

High-fidelity, small-scale patterning is often a tradeoff between full-pattern methods that may have limited resolution or flexiblity, and serial methods that can create high-resolution patterns but only by slow processes. Furthermore, metals have limited formability at very small scales. Gao *et al.* developed a method to create very smooth three-dimensional crystalline metallic nanoscale structures using a laser to create shockwave impulses. The shockwave creates ultrahigh-strain-rate deformations that overcome the metal's normal strength and, thus, resistance to patterning. — MSL

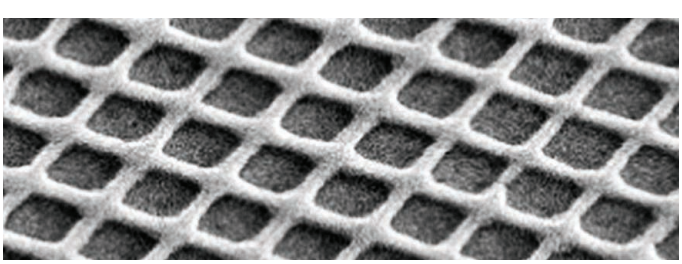
Science, this issue p. 1352

ONCOGENE REGULATION

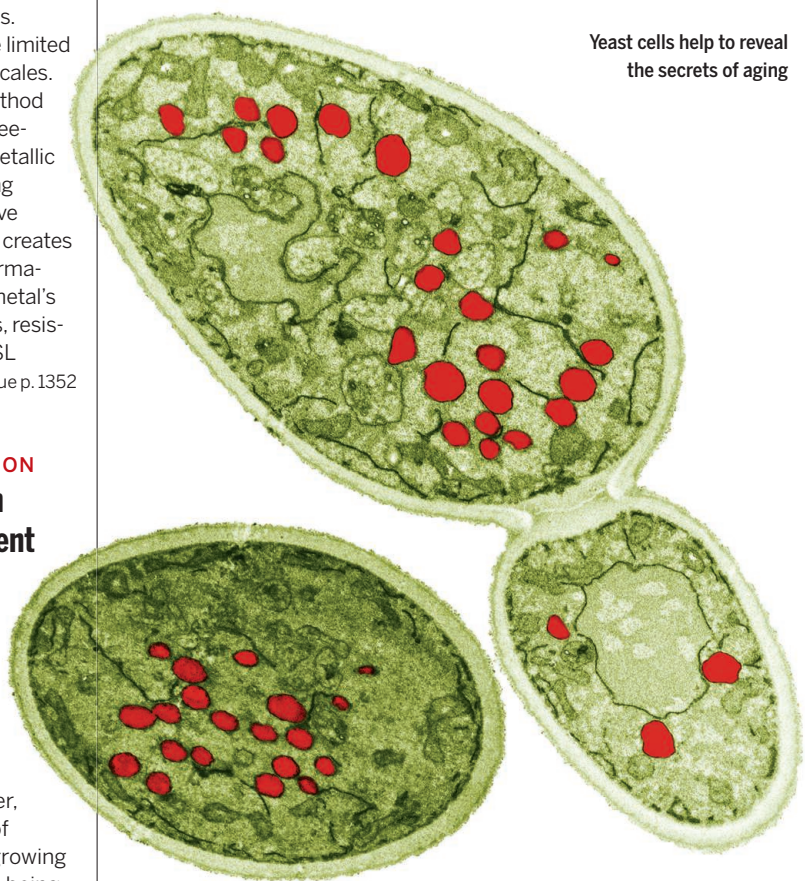
A super-enhancer in leukemia development

Human cancer genome projects have provided a wealth of information about mutations that reside within the coding regions of genes and drive tumor growth by functionally altering protein products. However, this mutational portrait of cancer is incomplete: A growing number of mutations are being found within gene regulatory regions. Mansour *et al.* present an intriguing example of this in a study of a childhood cancer, T-cell acute lymphoblastic leukemia (see the Perspective by Vähärautio and Taipale). An oncogene known to drive the growth of this cancer is expressed at high levels in the leukemic cells because the cells harbor mutations that create a powerful super-enhancer (a DNA sequence that activates transcription) upstream of the oncogene. — PAK

Science, this issue p. 1373; see also p. 1291



Silver fishnet structure from laser shock imprinting



Yeast cells help to reveal the secrets of aging

AGING IN YEAST

DNA circles involved in an aging SAGA

Budding yeast cells provide a powerful model system to study the mechanisms of aging. As they grow older, yeast cells accumulate extra-chromosomal ribosomal DNA circles, a by-product of DNA repair processes. Somehow, during cell division, these circles are retained in the mother cell where they replicate and limit mother cell viability. Denoth-Lippuner *et al.* report that the yeast acetyl-transferase complex SAGA plays a role in this process. Cells lacking SAGA are long-lived. SAGA promotes the retention of DNA circles within the nucleus of yeast mother cells by anchoring them to nuclear pores. The nuclear pores and DNA circles accumulate in a sort of cap inside the nuclear envelope, which is preferentially retained in the mother cell. — SMH

eLife **3**, e03790 (2014).

SOCIOLOGY

Networking doesn't always mean cooperating

The relationship between network structure and cooperation has been hard to define. Rand *et al.* arranged sets of volunteers in fixed networks, where they could only interact with specific players, in a Prisoner's dilemma game. In this game, "defectors" earned points and "cooperators" paid a cost so that each of their neighbors benefited. After each round, players learned what their neighbors had done. Not all network conformations

CREDITS: (TOP TO BOTTOM) BIOPHOTO ASSOCIATES/SCIENCE SOURCE; GAO *ET AL.*

Published by AAAS

ALSO IN *SCIENCE* JOURNALS

Edited by **Stella Hurtley**

SYSTEMS BIOLOGY

Dynamic signals enhance information transfer

Cells need to process information about their external environment reliably to survive. However, variation, or noise, in biochemical reactions, or in the states of individual cells, make it hard for a cell to detect concentration, rather than just the presence or absence of an activating ligand. Selimkhanov *et al*. show that cellular signaling circuits get around this problem by continually monitoring signals over time. Such dynamic responses in cultured human cells more effectively distinguish signals from noise and thus avoid loss of information transmitted to the cell from external signals. — LBR

Science, this issue p. 1370

AVIAN EVOLUTION

Bird evolution: How birds took to the air

Research on the origin and evolution of birds has gathered pace in recent years, aided by a continuous stream of new fossil finds as well as molecular phylogenies. Bird origins, in particular, are now better understood than those of mammals, for which the early fossil record is relatively poor compared with that of birds. Xu *et al.* review progress in tracing the origins of birds from theropod dinosaurs, focusing especially on recent fossil finds of feathered dinosaurs of northeastern China. They integrate current research on developmental biology and functional anatomy with the paleontological record, to show how key features of birds—feathers, wings, and flight—originated and evolved, and radiated from their dinosaur forebears. — SNV

Science, this issue p. 1341

SOLID-STATE THEORY

Predicting an exotic state of matter

Much like graphene, two-dimensional flakes of transition metal dichalcogenides have appealing electronic properties. Qian *et al*. now find that certain structures of these materials may also exhibit the so-called spin Hall effect. The spin Hall effect represents an exotic state of matter in which a 2D material conducts electricity along its edge in a way that drastically reduces dissipation. To show this, the researchers used first-principle calculations and found that the materials also feature a large band gap, which reduces undesirable conduction through the bulk. Their proposed device could be switched on and off quickly using an electric field. — JS

Science, this issue p. 1344

PLANETARY SCIENCE

Chemistry and physics of extraterrestrial life

Is there life on other planets? To answer this question, scientists must first determine what makes a planet habitable. Mars provides a natural laboratory to explore this question. In a Perspective, Conrad explains that, starting with the Viking landers in 1976, numerous missions to Mars have attempted to determine whether the planet could support life—or was able to do so in the past. Mobile rovers have explored the surface, aided by orbiters and, most recently, have been equipped with the tools to dig below the surface. By interrogating the chemical and physical conditions both now and in the past, these missions may soon be able to provide a more definitive answer. — JFU

Science, this issue p. 1288

ROCK MECHANICS

Nanofibers involved in fault rupture

Changing fault properties during rupture dictates the size and extent of an earthquake. Faulting leads to well-known microstructures that may play a role in how natural faults slip during rupture. Verberne *et al*. investigated tiny, nanogranular fibers found in microstructures generated on simulated carbonate faults. A microphysical model was able to account for how the small and aligned fiber produced runaway fault slip, similar to that seen in natural faults. These small structures play a role in carbonate faulting and similar microstructures could control fault rupture in other types of rocks. — BG.

Science, this issue p. 1342

OPHTHALMOLOGY

All eyes on human limbal stem cells

Our corneas—the transparent structures that allow us to see—are easily damaged by trauma and infection, which can cause scarring or blindness. Although corneas can be transplanted, transplants are limited by immune responses and by a shortage of cornea donors. Basu *et al.* devised a cell-based approach to prevent corneal scarring. They obtained stem cells from the human limbus (the region between cornea and sclera), which could be differentiated into keratocytes (corneal cells). The stem cells actively regenerated new corneal tissue when encased in a fibrin gel and applied to the wounded surface of the eye in mice. Such cells could potentially be obtained directly from a patient to treat scarring and prevent blindness. — MLF

Sci. Transl. Med. **6**, 266ra172 (2014).

ELECTRON MICROSCOPY

A golden era for electron microscopy

Electron microscopy (EM) is an attractive method for determining structures of protein complexes that are difficult to crystallize. Exciting recent developments in electron detectors allow EM structure determination to near atomic resolution. A key impediment to further improvement is that single specimens move during irradiation. Russo and Passmore designed a gold support that moves much less during irradiation than the current support and as a result prevents movement of the protein sample. Using the support they determined the structure of apo-ferritin which, as a spherical shell of α-helices, is particularly challenging to solve by EM. — VV

Science, this issue p. 1377

CANCER

A receptor tyrosine kinase signals to YAP

The Hippo pathway limits cell proliferation by inhibiting the activity of the transcriptional coactivator YAP. In contrast, cell proliferation is stimulated by the binding of growth factors to tyrosine receptor kinases, such as the binding of neuregulin to ERBB4. Neuregulin binding also triggers the cleavage of ERBB4. Haskins *et al.* found that a fragment containing the intracellular domain of ERBB4 interacted with and activated YAP (see the Perspective by Sudol). Breast cancer cell migration induced by neuregulin was blocked by knocking down YAP. Thus, ERBB4 could promote tumor aggressiveness both through receptor tyrosine kinase signaling and by stimulating YAP. — LKF

Sci. Signal. **7**, ra116 and pe29 (2014).

Published by AAAS



A coffee plant pathogen is leaving a trail through South America

PLANT PATHOLOGY

Morning brew at risk

Coffee leaf scorch has affected yields of coffee in Brazil, the supplier of much of the world's coffee, since the 1990s. The disease is caused by *Xylella fastidiosa*, a bacterium that clogs xylem vessels and impedes transport of water and nutrients for the plant, resulting in a degradation of the quality and yield of the coffee beans and the eventual death of the plant. Nunney *et al.* genotyped bacterial samples from infected plants in Costa Rica to reconstruct the pathogen's progress as it slithers from Brazil and through the Americas. Citrus groves are also at risk because substrains of this pathogen hopscotch into new domains, perhaps using homologous recombination with strains already present. — PJH

PLOS One **9**, e112463 (2014).

increased cooperation—the rewards relative to costs had to be greater than the number of linked players in order for cooperation to increase and clusters of cooperators to form. — BJ

Proc. Natl. Acad. Sci. U.S.A. 10.1073/pnas.1400406111 (2014).

WATER STRUCTURE

Getting all wet 9 or 10 at a time

H_2O seems like such a simple molecule. Why, then, haven't chemists definitively resolved the details of liquid water structure? The trouble is that hydrogen bonding—the attraction between an H on one molecule and an O on another—creates a cooperative net, where the orientation of each individual H_2O depends sensitively on the orientations of all its neighbors. Pérez *et al.* take a close look at water clusters built up from 9 or 10 molecules, to help tease out the subtleties of hydrogen bonding behavior in a more manageable size regime. Using rotational spectroscopy in the microwave region, they observe several different arrangements at slightly varying energies, which accord with a straightforward cooperativity model. — JSY

Angew. Chem. Int. Ed. **53**, 10.1002/anie.201407447 (2014).

ECONOMICS

Working from home can work well

In the U.S., telecommuting is most commonly observed for employees in the highest and lowest income deciles, yet there is little evidence about its benefits. Bloom *et al.* carried out the first randomized control trial of this management practice and report that performance of at-home workers increased by 13%. They found that Ctrip, a large, China-based travel agency, experienced lower turnover among its call-center workers, who also reported greater job satisfaction. When this program was rolled out to the entire company, some of the at-home employees elected to return to the office, whereas others switched to working from home; this reassortment improved the overall performance gain to 22%. — GJC

Quart. J. Econ. 10.1093/qje/qju032 (2014)

ICE SHEETS

Bottom-up effects on Northern glaciation

How much did the solid Earth contribute to the development of Northern Hemispheric glaciation around 3 million years ago? Much attention has been paid to the roles of solar insolation, distal orogeny, and the ocean, and now Steinberger *et al.* suggest that regional geodynamic processes also had a profound influence on the growth of major ice sheets. They argue that the mantle plume beneath Iceland thinned the lithosphere and caused regional uplift in Greenland, while Greenland was drifting northward due to plate tectonics and true polar wander. These factors could help explain the reappearance of major Northern Hemispheric ice sheets after an absence of more than 500 million years. — HJS

Terra Nova 10.1111/ter.12133 (2014)

ENVIRONMENTAL SCIENCE

A simple, efficient way to purify water

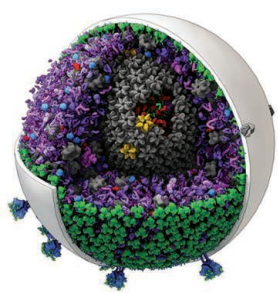
Unsafe drinking water is a serious public health issue that affects many millions of people worldwide. Improved methods for purifying water have the potential to save many lives in the developing world, but they have to be affordable. Ehdaie *et al.* have developed a prototype ceramic tablet that, when added to a container, can purify 10 liters of water each day for more than 154 days. The tablet works by releasing silver ions, which effectively destroy bacteria, viruses, and protozoa, but are safe for humans at the levels released. These devices are comparatively cheap and easy to produce from sawdust, clay, and silver nitrate. Differences in clay composition can affect the tablet properties, so the approach still must be tested under real-world conditions. — JFU

Environ. Sci. Technol. 10.1021/es503534c (2014).

STRUCTURAL BIOLOGY

HIV in 3D: Modeling the mesoscale

Structural biologists have one set of tools to visualize atomic-level details of large structures and another that reveals the nuances of small ones. But detailed models of medium-sized viruses, synaptic vessels, plasma, and the like—objects at the so-called mesoscale, ranging from 10 to 100 nanometers—have largely remained in the dark. Now, Johnson *et al.* have developed cellPACK, freely available software that produces 3D images of biological systems at the mesoscale. As a demonstration of cellPACK's utility, the tool purports to resolve a longstanding debate about whether HIV surface proteins are randomly distributed on the virus's surface (spoiler alert: they are not). — JC



cellPACK's HIV model shows protein distribution on the virus's surface.

Nat. Methods, 10.1038/nmeth.3204 (2014)

CREDITS: (TOP TO BOTTOM) GERSON SOBREIRA/TERRASTOCK; GRAHAM JOHNSON AND LUDOVIC AUTIN/THE SCRIPPS RESEARCH INSTITUTE

Published by AAAS

REVIEW SUMMARY

AVIAN EVOLUTION

An integrative approach to understanding bird origins

Xing Xu,* Zhonghe Zhou, Robert Dudley, Susan Mackem, Cheng-Ming Chuong, Gregory M. Erickson, David J. Varricchio

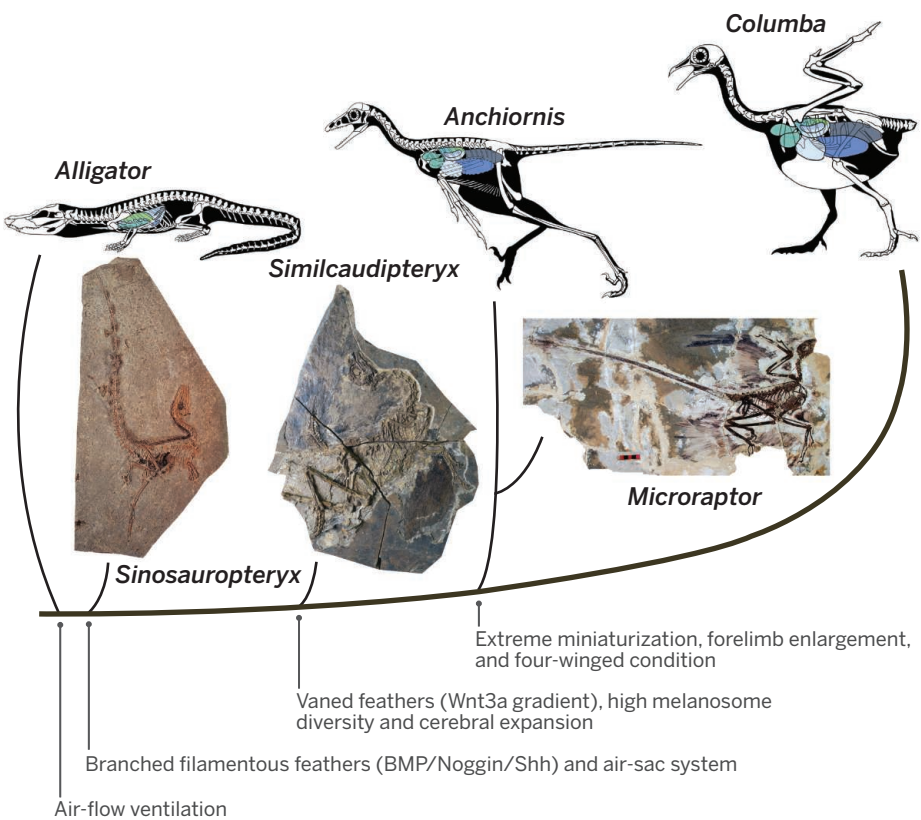
BACKGROUND: The origin of birds is one of the most enduring and dramatic evolutionary debates. The hypothesis that the primarily small-sized birds are nested within a theropod dinosaur group that includes the gigantic *Tyrannosaurus rex* has been supported by strong fossil evidence, but until recently, several important issues remained unresolved, including the origins of feathers and flight, the "temporal paradox" (the coelurosaurian theropods occur too late in the fossil record to be ancestral to the Jurassic bird *Archaeopteryx*), and supposed homological incongruities (e.g., the suggested homologies of three fingers in tetanuran theropods are different from those of living birds). Recent discoveries of spectacular dinosaur fossils from China and elsewhere provide new information to address these issues.

ADVANCES: The discoveries of feathered dinosaur fossils from the Jurassic and Cretaceous sediments of China and elsewhere document a diverse range of feathers from monofilamentous feathers to highly complex flight feathers, which show a general evolutionary trend of increasing complexity leading to the cladogenesis of birds. The wide occurrence of foot feathers in Mesozoic theropods (i.e., short filamentous forms in relatively basal theropods and large vaned forms in derived theropods, including early birds) clarifies feather-scale relations and integumentary evolution pertinent to flight origins and also shows that bird flight likely evolved through a four-winged stage. With numerous discoveries of well-preserved dinosaur fossils over a wide range of geological periods, the morphological, functional, and temporal transition from ground-living to flight-capable theropod dinosaurs is now one of the best-documented major evolutionary transitions. Meanwhile, studies in disciplines other than paleontology provide new insights into how bird characteristics originated and evolved—such as feathers, flight, endothermic physiology, unique strategies for reproduction and growth, and an unusual pulmonary system. The iconic features of extant birds, for the most part, evolved in a gradual and stepwise fashion throughout theropod evolution. However, new data highlight occasional bursts of morphological novelty at certain stages close to the origin of birds and an unavoidable complex, mosaic evolutionary distribution of major bird characteristics on the theropod tree. Research into bird origins provides a model of how an integration of paleontological and neontological data can be used to gain a comprehensive understanding of the complexity surrounding major evolutionary transitions and to set new research directions.

ON OUR WEB SITE

Read the full article at http://dx.doi.org/10.1126/science.1253293

OUTLOOK: A refined, more robust phylogeny will be imperative to move our studies forward. A larger data set will help to increase the accuracy of phylogenetic reconstructions, but better character formulation and more accurate scorings are imperative at the current stage. In terms of character evolution, an integrative approach combining paleontological, neontological, developmental, temporal, and even paleoenvironmental data is particularly desirable. Greater examination of fossils pertaining to molecular information is also a potentially fruitful avenue for future investigation. Evolutionary scenarios for various aspects of the origin of birds have sometimes been constructed from neontological data, but any historical reconstruction must ultimately be tested using the fossil record. Consequently, dense fossil sampling along the line to modern birds and better understanding of transitional forms play key roles in such reconstructions. ■



Evolutionary history of selected bird features inferred from multidisciplinary data. Recent studies demonstrate that major bird characteristics have evolved in a sequential way, and many of them initiated transformation early in dinosaur evolution, with some approaching modern conditions well before the origin of birds, whereas others appear only near the origin of the crown group birds.

The list of author affiliations is available in the full article online.
* Corresponding author. E-mail: xu.xing@ivpp.ac.cn
Cite this article as X. Xu *et al.*, *Science* 346, 1253293 (2014).
DOI: 10.1126/science.1253293

Downloaded from www.sciencemag.org on December 14, 2014

Published by AAAS

REVIEW

An integrative approach to understanding bird origins

Xing Xu,[1]* Zhonghe Zhou,[1] Robert Dudley,[2] Susan Mackem,[3] Cheng-Ming Chuong,[4,7] Gregory M. Erickson,[5] David J. Varricchio[6]

Recent discoveries of spectacular dinosaur fossils overwhelmingly support the hypothesis that birds are descended from maniraptoran theropod dinosaurs, and furthermore, demonstrate that distinctive bird characteristics such as feathers, flight, endothermic physiology, unique strategies for reproduction and growth, and a novel pulmonary system originated among Mesozoic terrestrial dinosaurs. The transition from ground-living to flight-capable theropod dinosaurs now probably represents one of the best-documented major evolutionary transitions in life history. Recent studies in developmental biology and other disciplines provide additional insights into how bird characteristics originated and evolved. The iconic features of extant birds for the most part evolved in a gradual and stepwise fashion throughout archosaur evolution. However, new data also highlight occasional bursts of morphological novelty at certain stages particularly close to the origin of birds and an unavoidable complex, mosaic evolutionary distribution of major bird characteristics on the theropod tree. Research into bird origins provides a premier example of how paleontological and neontological data can interact to reveal the complexity of major innovations, to answer key evolutionary questions, and to lead to new research directions. A better understanding of bird origins requires multifaceted and integrative approaches, yet fossils necessarily provide the final test of any evolutionary model.

The origin of birds is one of the most enduring and dramatic evolutionary debates (*1*–*3*). Whether the primarily small-sized birds are nested within a theropod group including the gigantic *Tyrannosaurus rex* is of great interest to both the academic community and general public. More challenging is whether the origin of major bird characteristics can be traced among the Mesozoic terrestrial theropod dinosaurs. Recent discoveries of spectacular dinosaur fossils from China and elsewhere (Fig. 1) provide significant new information to address these issues and also prompt numerous studies in disciplines other than paleontology to explain how bird characteristics originated and evolved (*4*–*11*). Whereas these new discoveries and studies have addressed many issues on the origin of birds, they also open new research directions. A better understanding of the origin of birds requires a more integrative approach, and this is particularly true given that understanding bird origins is more about reconstructing the evolutionary patterns of major bird characteristics than just filling the gaps between birds and their theropod ancestors. In this review, we present a brief introduction to recent advances on this topic, discuss several issues that have interested both paleontologists and scientists in other fields, and finally provide our thoughts on how to proceed using an integrative approach in evolutionary biology. For the convenience of further discussion, we define the vernacular term "birds" (avians in many publications) to correspond to Avialae, a stem-based taxon defined as the most-inclusive clade containing *Passer domesticus* but not *Dromaeosaurus albertensis* or *Troodon formosus*.

[1]Key Laboratory of Vertebrate Evolution and Human Origins, Institute of Vertebrate Paleontology and Paleoanthropology, Chinese Academy of Sciences, Beijing, 100044, PR China. [2]Department of Integrative Biology, University of California, Berkeley, CA 94720, USA. [3]Cancer and Developmental Biology Laboratory, Center for Cancer Research, NCI-Frederick NIH, Frederick, MD 21702, USA. [4]Department of Pathology, University of Southern California, CA 90033, USA. [5]Department of Biological Science, Florida State University, Tallahassee, FL 32306-4295, USA. [6]Earth Sciences, Montana State University, Bozeman, MT 59717, USA. [7]Cheng Kung University, Laboratory for Wound Repair and Regeneration, Graduated Institute of Clinical Medicine, Tainan, 70101, Taiwan.
*Corresponding author. E-mail: xu.xing@ivpp.ac.cn

Family tree of birds and relatives

A reliable phylogeny sets the framework for reconstructing the evolutionary sequence of major bird characters. Although birds are widely accepted to form a monophyletic group (i.e., a group composed of an ancestor and all its descendants) within the Archosauromorpha, the identity of their closest relatives has been hotly debated (*1*–*3*, *12*).

The last four decades have witnessed an unprecedented accumulation of evidence supporting the hypothesis that birds are maniraptoran theropods (BMT hypothesis) (*2*, *3*, *13*), and most likely closely related to the troodontids and/or dromaeosaurs (*14*–*23*). The BMT hypothesis is supported by numerous skeletal, behavioral, and physiological resemblances between nonavialan theropod dinosaurs (hereafter, theropods) and birds. These resemblances are derived from such birdlike theropods as the dromaeosaur *Deinonychus*, the troodontid *Saurornithoides*, and others (*24*, *25*), dinosaur fossils preserving gastroliths for digestion (*26*), highly pneumatic theropod skeletons (*27*–*29*), dinosaur nests preserving brooding and parental care behavior (*30*–*33*), theropod eggs with distinctive bird features (*34*–*36*), and bone microstructures indicating fast growth rates (*37*), among others. Nearly all published numerical phylogenetic analyses support the BMT hypothesis (*15*–*23*, *38*). Comparatively, other hypotheses on bird origin are based on a small number of similarities from certain parts of the body between birds and the proposed groups (*1*). One recently published numerical phylogenetic analysis places birds and other maniraptoran theropods at the base of the Archosauria rather than within the Theropoda (*12*), but this analysis is flawed in several aspects including a strongly biased data set used for the analysis (*39*).

Although birds are now widely accepted to be theropods, maniraptorans, and paravians, there is an unresolved debate on what are the most basal birds. Suggested taxa include the probably flighted *Archaeopteryx* and *Rahonavis* (*40*), the enigmatic scansoriopterygids (*41*), the four-winged *Anchiornis* and its kin (*21*, *42*), and the Gondwanan unenlagiid theropods (*38*, *43*). Many recent studies suggest that *Rahonavis*, other unenlagiids, and *Anchiornis* and its kin are deinonychosaurs (*16*, *17*, *22*, *44*, *45*). Several analyses even place the iconic *Archaeopteryx* at the base of the Deinonychosauria (*22*), yet other studies suggest that the Deinonychosauria itself is paraphyletic, with the troodontids more closely related to birds than the dromaeosaurs (*21*) or vice versa (*20*). A radical suggestion is that the oviraptorosaurs are closely related to the scansoriopterygids (*38*, *46*, *47*) and, further, that they together represent the first major branching in bird evolution (*46*). Despite these debates, considerable consensus exists for the general branching pattern of major theropod groups (*15*–*23*, *38*), here used as a framework for further discussions on bird origins (Fig. 2).

Recently discovered transitional forms toward modern birds

Numerous recently discovered fossils of theropods and early birds have filled the morphological, functional, and temporal gaps along the line to modern birds. These discoveries have helped address several important issues repeatedly raised to challenge the BMT hypothesis (*1*, *48*, *49*), such as the origins of feathers and flight, the "temporal paradox" in the stratigraphic distribution of theropod fossils (i.e., most of the diversity of coelurosaurian theropods occurs much later in the fossil record than *Archaeopteryx* in the Late Jurassic), and the supposed homological incongruities (e.g., the suggested homologies of three fingers in tetanuran theropods are different from those of living birds).

Numerous discoveries demonstrate that many bird characteristics have their origins among theropods or a more inclusive group (Fig. 2). For example, the discoveries of *Mahakala* and other basal members of derived theropod groups

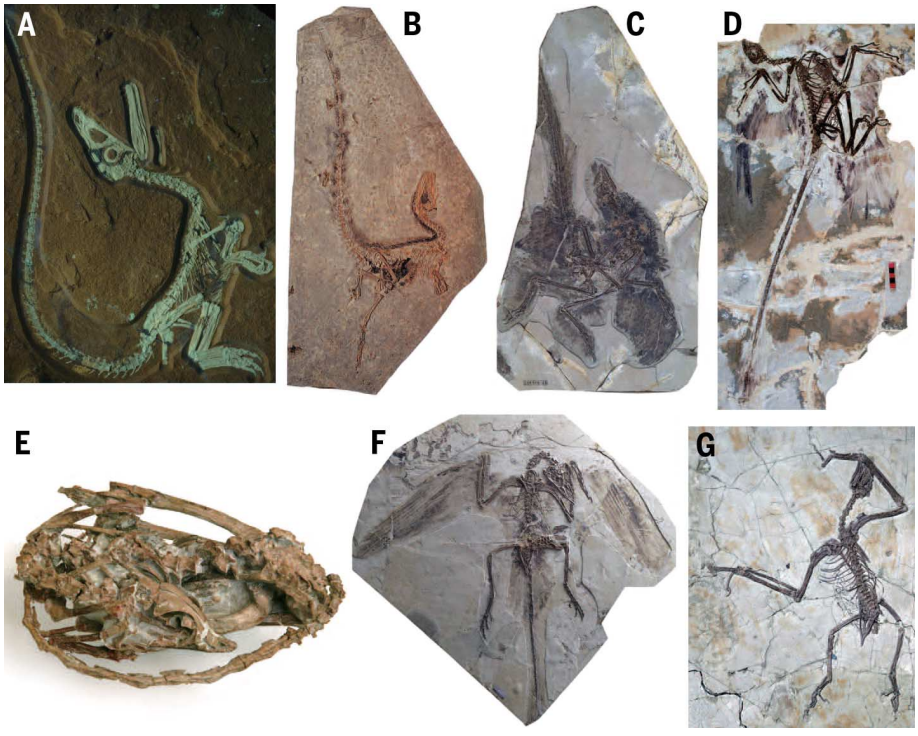


Fig. 1. Selected recently discovered transitional forms of theropods and early birds. The basal tetanuran *Sciurumimus* (**A**) and the basal coelurosaur *Sinosauropteryx* (**B**) with filamentous feathers. The deinonychosaurs *Anchiornis* (**C**) and *Microraptor* (**D**) with numerous skeletal and integumentary features including large vaned feathers on hindlimbs that are also present in basal birds. (**E**) The troodontid *Mei* with a birdlike tuck-in sleeping posture. The basal birds *Jeholornis* with a dromaeosaur-like tail (**F**) and *Sapeornis* with a typical deinonychosaurian shoulder girdle (**G**).

suggest that extreme miniaturization and laterally movable arms necessary for flapping flight are ancestral for paravian theropods (*43*, *50*, *51*). The ceratosaur *Limusaurus*, with a highly reduced pollex, might represent a key stage in theropod hand evolution (*52*); the megalosaur *Sciurumimus* (Fig. 1A), the compsognathid *Sinosauropteryx* (Fig. 1B), and a few other dinosaurs—including even some ornithischians—document the appearance of primitive feathers (*53*–*57*). Other discoveries include the oviraptorosaur *Caudipteryx* (*58*), the deinonychosaur *Anchiornis* (Fig. 1C), *Microraptor* (Fig. 1D), and several other maniraptorans that have pennaceous (vaned) feathers (*45*, *57*–*59*). The troodontid *Mei* (Fig. 1E) preserves a birdlike "tuck-in" sleeping posture (*60*). In contrast, a number of basal birds resemble theropods in many features. For example, new *Archaeopteryx* specimens indicate that this early bird has tetraradiate palatine, as in theropods, and probably a specialized second pedal digit as in deinonychosaurs (*61*); *Jeholornis* (Fig. 1F) has a dromaeosaur-like long, stiff bony tail (*62*); and the short-tailed *Sapeornis* (Fig. 1G) has a shoulder girdle similar to those of the deinonychosaurs (*63*).

Earlier studies demonstrate that the temporal paradox is not a real problem for the BMT hypothesis, which is in fact more consistent with the stratigraphic record than any other alternative phylogenetic hypothesis regarding the origin of birds (*64*). Nonetheless, recent investigations in the Jurassic have resulted in the discoveries of well-preserved fossils of several derived theropod groups, including Tyrannosauroidea (*65*), Alvarezsauroidea (*23*), Troodontidae (*45*), and possibly Therizinosauroidea (*66*) and Oviraptorosauria (*38*, *41*). These discoveries improve the stratigraphic fit of the theropod fossil record to the BMT hypothesis (*23*) and have effectively resolved the so-called temporal paradox problem by showing that many coelurosaurian groups occur earlier than *Archaeopteryx* in the fossil record (Fig. 2).

Newly discovered fossils and relevant analyses demonstrate that salient bird characteristics have a sequential and stepwise transformational pattern, with many arising early in dinosaur evolution, undergoing modifications through theropods, and finally approaching the modern condition close to the origin of the crown group birds (Fig. 2). For example, the unusually crouched hindlimb for bipedal locomotion that characterizes modern birds was acquired in stepwise fashion through much of theropod evolution (*67*), and both the furcula (*68*) and the "semilunate" carpal (*69*) appeared early in theropod evolution. Notably, major bird characteristics often exhibit a complex, mosaic evolutionary distribution throughout the theropod tree, and several evolutionary stages are characterized by accelerated changes (*70*). For example, the early evolution of paravian theropods features cerebral expansion and elaboration of visually associated brain regions (*71*), forelimb enlargement (*22*, *67*), acquisition of a crouched, knee-based hindlimb locomotor system (*67*), and complex pinnate feathers associated with increased melanosome diversity, which implies a key physiological shift (*72*). Together these features may suggest the appearance of flight capability at the base of the Paraves (*22*, *67*).

Transitional features in dinosaurian reproduction and growth

Living birds display a set of unique reproductive features unknown in other extant vertebrates. Nevertheless, based on the study of eggs, eggshell, embryos, and nesting traces, most paleontologists agree that various anatomical features including aspects of egg shape, ornamentation, microstructure, and porosity characteristic of living birds trace their origin to the maniraptoran theropods, such as oviraptorosaurs and troodontids. Currently, our understanding of reproduction in dromaeosaurs remains limited, as only one taxon is associated with any reproductive material, namely, a *Deinonychus* specimen found on top of eggs (*33*). These features, together with adult-egg associations, clutch configurations, and nest traces, further suggest that associated behavioral and physiological attributes may also have evolved before the origin of birds. These maniraptoran dinosaurs share with modern birds: eggs with hard calcitic shells with narrow units relative to overall shell thickness, a second structural layer of vertical prisms, sparse and narrow pores, at least some squamatic structure, and calcium absorption ("cratering") of the mammillae by the developing embryo (*34*–*36*, *73*). Egg size relative to adult size in both oviraptorosaurs and troodontids is large in comparison to the ratios for all other theropods (*74*). Further, these eggs greatly exceed those typical of modern reptiles of similar body mass but are less than 40% the size expected in a bird of similar adult body mass (*74*).

On the basis of discoveries of within-clutch egg pairing (*31*) and of specimens with two eggs within or in close proximity to adult skeletons (*75*, *76*), maniraptorans have been inferred to have sequential ovulation, as in birds, but from two functional ovaries and oviducts (monoautochronic ovulation) (*31*). Adult-clutch associations, some preserving brooding postures, are known for four oviraptorosaurs (*30*, *77*), two troodontids (*31*, *32*), a dromaeosaur (*33*), and one basal bird (*74*) providing clear evidence for parental care of eggs. The extra-large clutches of these dinosaurs exceed the mass of those for equivalent-sized extant reptiles and birds by two to fourfold (*78*). Such large clutch sizes are akin to those of modern birds showing paternal or maternal parental care, as opposed to those exhibiting biparental care (*79*). None of the clutch-associated adults (n = 8) of nonavialan maniraptorans has any osteo-histologic features associated with egg-laying females. Thus, the large clutches of oviraptorosaurs and troodontids may favor communal nesting and paternal (male-only) care for these groups (*80*, *81*).

The eggshell and eggs of oviraptorosaurs and dromaeosaurs, where known, are two-layered, moderately porous, and only weakly asymmetric (*33*, *34*, *36*, *73*), although the ornamented eggs of alvarezsaurs have a three-layered eggshell (*82*). In contrast, troodontids exhibit a more strongly asymmetric egg lacking ornamentation with potentially three structural shell layers and low porosity (Fig. 3A), and they are structured more like modern bird eggs than like those of other dinosaurs (*83*, *84*). In addition, the tighter clutch configuration, greater exposure of eggs (Fig. 3B), and avialan levels of porosity favor contact incubation in troodontids (*85*, *86*).

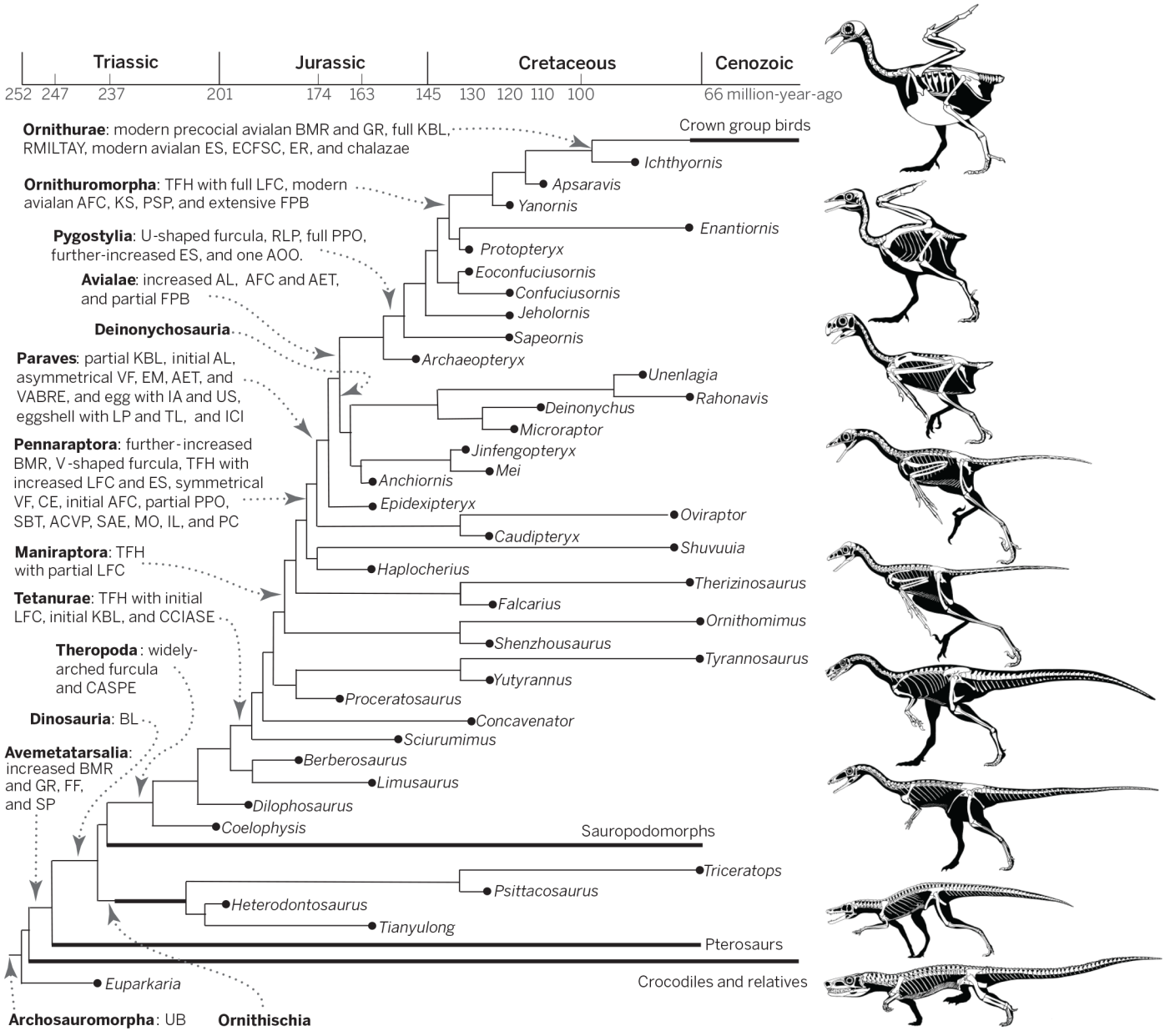


Fig. 2. Selected species on a temporally calibrated archosauromorphan phylogeny showing the evolution of major characteristics along the bird-line. The phylogeny is a combination of two recently published analyses: The basal part of the tree is derived from (*55*) and the upper part from (*16*). Skeletal silhouettes of several archosauromorphans show the general morphological features along the bird-line within this group, and they are the basal archosauromorphan *Euparkeria*, the basal crocodilomorph *Sphenosuchus*, the basal theropod *Coelophysis*, the basal coelurosaur *Protoceratosaurus*, the basal paravian *Anchiornis*, the basal avialan *Archaeopteryx*, the basal pygostylian *Sapeornis*, the basal ornithuromorphan *Yanornis*, and the crown group bird *Columba* (bottom to top). Acronyms in the figure: ACVP, advanced costosternal ventilator pump; AET, arm elongation and thickening; AFC, arm flapping capability; AL, aerial locomotion; AOO, active ovary and oviduct; BL, bipedal locomotion; BMR, basal metabolic rate; CASPE, cervical air sacs posterior extension; CCIASE, cranial and caudal intrathoracic air ascs elaboration; CE, cerebral expansion; ECFSC, egg clutch free of sediment cover; EM, extreme miniaturization; ER, egg rotation; ES, egg size; FF, filamentous feathers; FPB, fusion of pelvic bones; GR, growth rate; IA, increased asymmetry; ICI, improved contact incubation; IL, iterative laying; KBL, knee-based locomotion; KS, kinetic skull; LFC, laterally folding capability; LP, low porosity; MO, monoautochronic ovulation; PC, paternal care; PPO, pubis posterior orientation; PSP, plowshare-shaped pygostyle; RLP, rodlike pygostyle; RMILTAY, rapid maturity in less than a year; SAE, slightly asymmetrical egg; SBT, short bony tail; SP, skeletal pneumatization; TFH, three-fingered hand; TL, third (external) layer; UB, unidirectional breathing; US, unornamented surface; VABRE, visually associated brain regions elaboration; VF, vaned feathers.

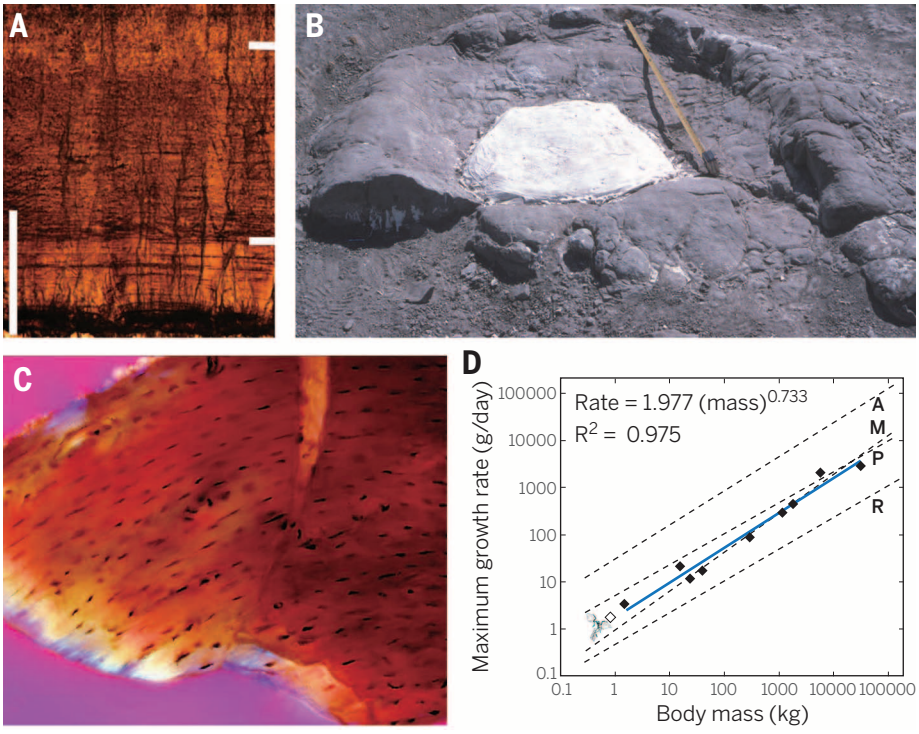


Fig. 3. Maniraptoran reproduction and growth. (**A**) Eggshell of *Troodon* in radial thin section showing three microstructural layers (separated by white bars) as common to many extant birds. Scale bar, 0.5 mm. (**B**) *Troodon* nesting trace with clutch of 24 eggs preserved under white plaster jacket. Tops of the eggs were exposed within the nest structure and were incubating by an attending adult representing a transitional stage between the fully buried clutches of most nonavialan dinosaurs and the sediment-free clutches typical of modern birds. Tape measure equals 1 m. (**C**) Histological section of an *Archaeopteryx* femur. The section, shown in polarized light microscopy, shows the characteristic long bone histology of basal-most birds composed of parallel fibered bone matrix with negligible vascularization. Growth lines are also commonplace but are not evident in this particular specimen (Munich *Archaeopteryx*: Bayerische Staatssammlung für Palaontologie und Geologie BSPG 1999/50). The histology matches that of same-sized nonavialan dinosaurs but not that of modern birds. (**D**) Maximal growth rates for *Archaeopteryx* compared with nonavialan dinosaurs and living vertebrates. *Archaeopteryx* growth rates (skeleton) fall on the line at the lowermost bound of the nonavialan dinosaur regression line (blue; individual species values are shown as black diamonds). R, typical extant reptilian growth rates, P, typical extant precocial land bird growth rates, M, typical extant marsupial growth rates, A, typical extant altricial land bird growth rates. Graphic is from (*98*).

Despite these similarities with modern birds, oviraptorosaurs, troodontids and even Cretaceous enantiornithine birds differ reproductively from modern birds in a few key features, including more elongate egg shape and eggs largely buried with sediment during incubation (*73*, *74*, *87*, *88*). Relative egg size increases markedly from theropods to basal birds of the Mesozoic, with the latter laying eggs 50 to 70% the expected size of those in extant birds (*74*). Several specimens of basal birds preserve a single mass of circular objects. Although their interpretation remains debated (*89*), these may represent ovaries with mature ovarian follicles and suggest that basal birds had only one active ovary and oviduct as in most modern birds (*90*). Modern birds differ from enantiornithines in producing relatively larger eggs that are incubated in the absence of sediment (*74*). These features highlight the gradual acquisition of "modern" reproductive traits from Mesozoic theropods, through Mesozoic birds such as enantiornithines, and on to modern birds.

Reproductive adaptations may have implications for three other aspects of bird evolution: the development of feathers, metabolism, and flight. First, elongate forelimb and tail feathers in oviraptorosaurs parallel the brooding postures adopted by adults, and lengthening of the feathers may reflect selection for egg care (*91*). Reproduction in maniraptoran dinosaurs may also imply physiological similarities with birds: Egg pairs suggest a reproductive output similar to that of basal birds (*74*). Additionally, iterative laying, together with the complex clutch configurations and presumed synchronous hatching (*83*) in these theropods, further implies that both adult body and incubation temperatures were elevated over ambient conditions (*92*). Finally, ground nesting, the basal condition in modern birds, has been used to argue for a "ground-up" origin of flight (*11*). However, other factors may have initially constrained ground nesting. Through the theropod-bird transition, eggs were incubated within sediments, which suggests that they may have lacked chalazae (*31*, *74*) [chords of albumen present in birds that allow the embryo to maintain an upright position during egg rotation (*93*)]. Thus, without chalazae, arboreal maniraptoran dinosaurs may have been required to incubate eggs within sediments in order to prevent detrimental egg rotation (*31*), as occurs in modern crocodilians (*94*).

The development, growth, and physiology of theropods and basal birds have been explored using a variety of approaches. Birds are notable among living vertebrates in being endotherms, developing explosively to adult sizes in less than a year (very large forms such as ostriches are exceptions), and having exceptionally high basal metabolic rates (*95*). However, some osteohistological studies show that the bone histology of enantiornithine birds is nearly avascular, composed of slowly deposited parallel-fibered matrix, and having growth lines (*96*), which are unlike the bones of most dinosaurs (i.e., with fast-growing and highly vascularized fibro-lamellar bone with growth lines), and these studies suggest that basal birds had slowed their growth compared with their dinosaur ancestors. Furthermore, because maximal growth rates strongly correlate with metabolic rates (*95*), they likely had lower basal metabolic rates than dinosaurs.

Construction of mass-age growth curves showed that nonavialan dinosaurs as a whole, although probably endothermic, did not have a physiology exactly like modern birds (*97*). They grew faster than living reptiles but somewhat slower than most precocial birds and considerably slower than altricial forms scaled to the same size. Histologic analysis shows that several derived theropods, *Archaeopteryx*, and other basal birds inherited this dinosaurian physiology and took over a year to reach somatic maturity (*98*) (Fig. 3, C and D). The last conclusion has been supported by further growth line evidence in other basal birds (*99*). The miniaturization event, once associated with the cladogenesis of birds, actually occurred in paravian theropods (*50*, *51*, *98*), and consequently, the nearly avascular bone in some basal lineages of birds and small theropods is simply a scaling effect. Modern bone types and, presumably, the remarkable elevated metabolic rates of birds—which allow growth to adult sizes in less than a year—appear to have evolved near the cladogenesis of ornithurine birds during the Cretaceous Period (*37*, *98*, *100*), although increased growth and metabolic rates occurred early in theropod evolution (Fig. 2). [Note: The pioneering studies that established dinosaur growth rates were recently criticized for insufficient sample size and other issues (*101*). However, a subsequent comprehensive specimen-rich analysis (*102*) strongly supports both the methodology and conclusions about dinosaur and basal bird growth rates and physiology.]

Complex evolution of avialan traits

Bird morphology features many unique adaptations, and the evolution of some of these traits is so complex that a comprehensive understanding is wanting. How the theropod hand evolved

into the bird wing continues to be a hotly debated issue in evolutionary biology. Traditionally, paleontological analysis of theropod digit reductions indicated that the three-digit hand arose by unilateral loss of digits IV and V, resulting in a I, II, III digit formula (*103*). In contrast, the use of "positional" criteria of embryologic precursors to identify digits in modern birds revealed a digit 2, 3, 4 formula (*104*, *105*). To resolve this paradox, it has been proposed that a homeotic transformation of digits occurred that converted the embryonic precursors to I, II, and III in the adult (frameshift hypothesis) (*103*). Confounding any reconciliation of evidence from these disciplines, paleontological analyses typically rely on adult morphology to infer evolutionary links, whereas developmental studies generally focus on very transient, vestigial embryonic remnants to infer homologies with adult structures that have been lost entirely during evolution. In contrast to cellular fate determination, morphological identity of a complex structure is generally not defined by any specific and unique genetic "markers," which creates difficulty in assigning unique homologies.

Nevertheless, the analysis of expression of spatially regulated genes in the embryonic limb, which can be considered "phenotypic" criteria for digit assignment, has provided support for the frameshift hypothesis, in that certain gene expression patterns normally correlated with digit I are retained in the first digit of the chick embryo hand. In addition, recent, more comprehensive genome-wide expression profiling comparing chick embryo hand and foot digit precursors also supports a I, II, III phenotypic formula, in that the expression profiles of the first digit in both the hand and foot are remarkably similar (*106*). Genetic lineage tracing of *Shh*-expressing cells in the chick, which normally contribute to digits IV and V in the pentadactyl mouse, further suggests that digit IV is phenotypically absent from the chick hand (*8*, *107*). All of the phenotypic data from chick embryos are likewise compatible with the classic paleontological interpretation of a digit I-II-III formula, if the avialan embryo "positional" evidence of digit precursors is discounted as being either misleading or incorrect. However, this requires a shift in the primary limb axis to run through digit 3 (axis shift hypothesis), both in extinct and modern tetanuran theropods (*108*), rather than through digit 4, as occurs in most tetrapods except urodeles (*109*), because the primary limb axis must, by definition, include the first digit to form in the embryo. In fact, a shortcoming of the positional data on vestigial digit condensations appearing in the embryonic hands of various avialan species is that, although a remnant of a positional fourth condensation (which could be either IV or V) is clearly posterior (lateral) to the three digits that are retained in the adult, the identification of a vestigial anterior (medial) condensation is controversial, and it is based on indistinct elements that could represent other structures [see review in (*110*)]. A reevaluation of digit condensation formation in mouse embryos also suggests that some of the traditional approaches used to visualize early digit condensations may be inadequate and could miss evidence for noncontiguous digit loss (gaps between condensations). In mouse, genetic modulation of signals critical for regulating digit pattern and number across vertebrate species results in the initial loss of a central, rather than lateral, digit (*111*, *112*). This suggests that, in principle, a I-II-IV formula is also possible (central loss hypothesis). The I-II-IV formula would explain the persistence of digit 1 and still retain the primary axis extending through digit 4, which is the accepted rule in all living tetrapods except urodeles. Note that the expression profiling data in chick (*106*) also revealed mixed expression features for the lateral digits, which could be construed as compatible with a I-II/III-IV avialan hand formula (or the presence of digit IV). Taken together, these developmental studies now leave open the possibility of I-II-III, or even I-II-IV or I-III-IV formulas, for the modern avialan hand. Conversely, more recent analyses of digit morphologies in ceratosaurian and basal tetanuran theropods (*52*) lend some support to a II, III, IV theropod formula with secondary transformations to produce final I, II, III morphologies through multiple, partial homeotic transformations (lateral shift hypothesis).

To reconcile the paradox of digit homologies between theropods and birds, all of the proposed models (Fig. 4A) require either multiple events entailing the reduction and then reappearance of a positional digit 4 to retain the primary limb axis or require that the primary axis has undergone a novel shift in its position from digit 4 to 3. Clearly, whereas much progress has been made, a final resolution to this ongoing debate demands additional work from both paleontological and developmental perspectives (*110*).

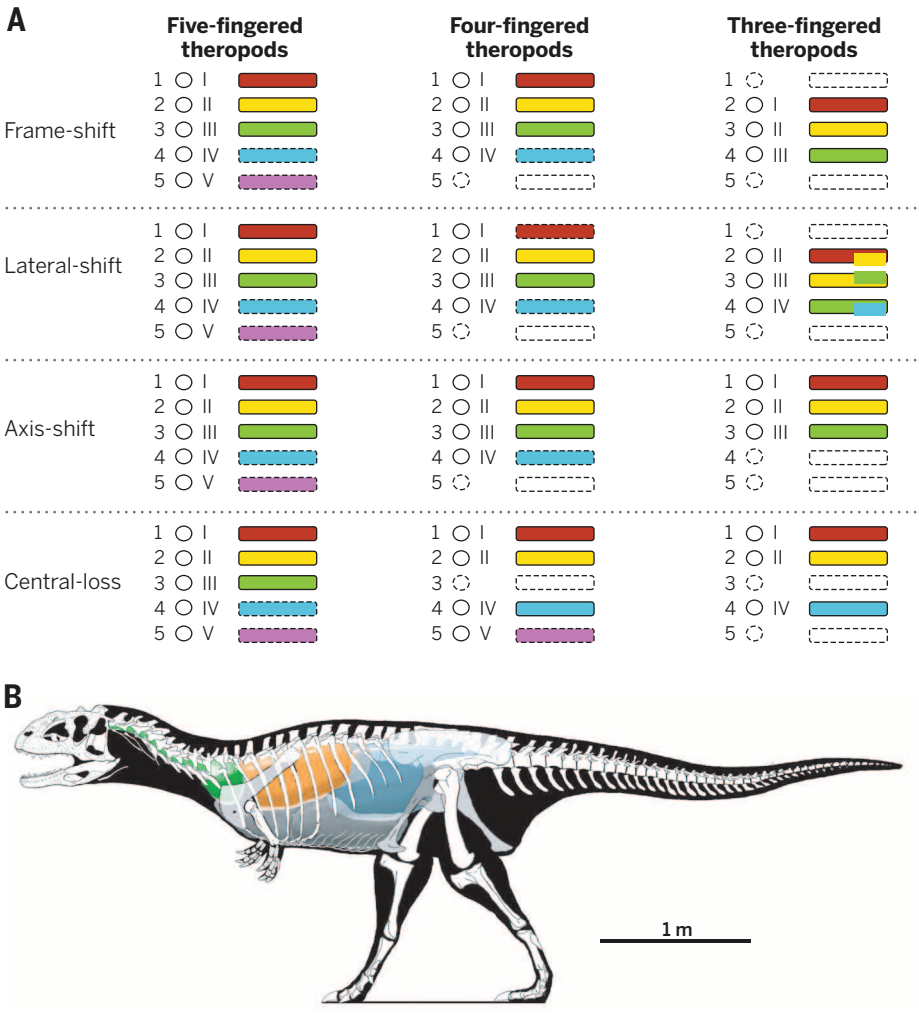


Fig. 4. The evolution of theropod hand and respiratory system. (**A**) Diagrams depicting four hypotheses for the evolution of wing digits. Color shading bordered by solid line refers to fully formed digits and by dash line to reduced digits. Colored bars refer to different "morphologic" digits. Arabic numerals refer to positional embryonic digit precursors. A combination of different colors refers to a mixture of different morphological features pertaining to different phenotypes in one digit. (**B**) Reconstruction of pulmonary components [cervical air-sac system (green), lung (orange), and abdominal air-sac system (blue)] in the theropod *Majungatholus*. Graphic from (*27*).

A highly efficient flow-through ventilation and a high-compliance air-sac system uniquely characterize the avialan respiratory system (*27*). Perceived differences in the respiratory systems of theropods and other reptiles, such as crocodilians, from those of birds have been invalidly used as evidence against the BMT hypothesis (*49*). Nevertheless, cervical and anterior dorsal vertebrae are clearly pneumatic in nearly all theropods (*29*), and documentation of extensive skeletal pneumaticity in theropods such as *Majungasaurus* and *Aerosteon* (*27*, *28*) demonstrates that a complex air-sac system and birdlike respiration evolved in birds' theropod ancestors (Fig. 4B). Furthermore, recent anatomical and physiological studies reveal that alligators (*7*, *113*), and even monitor lizards (*114*), exhibit respiratory systems and unidirectional breathing akin to those of birds. These studies demonstrate that unidirectional breathing is a primitive characteristic of archosaurs or an even more inclusive group with the complex air-sac system evolving later within Archosauria. Weight savings might have driven the evolution of postcranial skeletal pneumatization in early theropods (*29*), and in addition to large-bodied theropods, small maniraptoran theropods such as basal oviraptorosaurs, dromaeosaurs, and troodontids evolved the complex air-sac system similar to that of birds (*28*, *29*).

Morphogenesis and early evolution of feathers

Feathers were once considered to be unique avialan structures. The presence of feathers is inferred in several theropods based on osteological correlates, e.g., quill knobs (*115*, *116*) [although the presence of quill knobs in the basal tetanuran *Concavenator* has been questioned (*117*)] or pygostyles (*118*). Short filaments have been identified as feathers in the alvarezsaur *Shuvuuia* (*119*). Direct evidence came to light in China in 1996 with the discovery of the feathered *Sinosauropteryx* (*120*). Since then, numerous specimens of most theropod groups and even three ornithischian groups preserving feathers have been recovered from the Jurassic and Cretaceous beds of northeastern China (*13*, *21*, *41*, *45*, *57*) and from the Jurassic and Cretaceous beds of Germany, Russia, and Canada (*55*, *56*, *121*–*123*). These feathers fall into several major morphotypes (Fig. 5), ranging from monofilamentous feathers to highly complex flight feathers (*57*). The phylogenetic distribution of these feathers largely concurs with the predictions from a developmental model of living birds (*4*), although some of these morphotypes are absent in living birds (*57*, *124*).

The findings of feathers in dinosaurs prompted developmental studies over the last decade investigating the molecular mechanisms regulating feather development and regeneration in chickens (*125*). We can now propose the following "molecular circuits" regarding tissue morphogenesis and the evolution of modern feathers. Each molecular circuit has specific features and can be modulated to produce a diversity of morphology:

(i) Cylindrical topology and the downshift of stem cell ring. The formation of loose mesenchyme in the bud core allows a tube configuration, with basal layer facing inside. Driven by Wnt signaling, feather stem cells now are positioned near the base, forming a ring configuration (*126*, *127*). It can remain as a tube or the

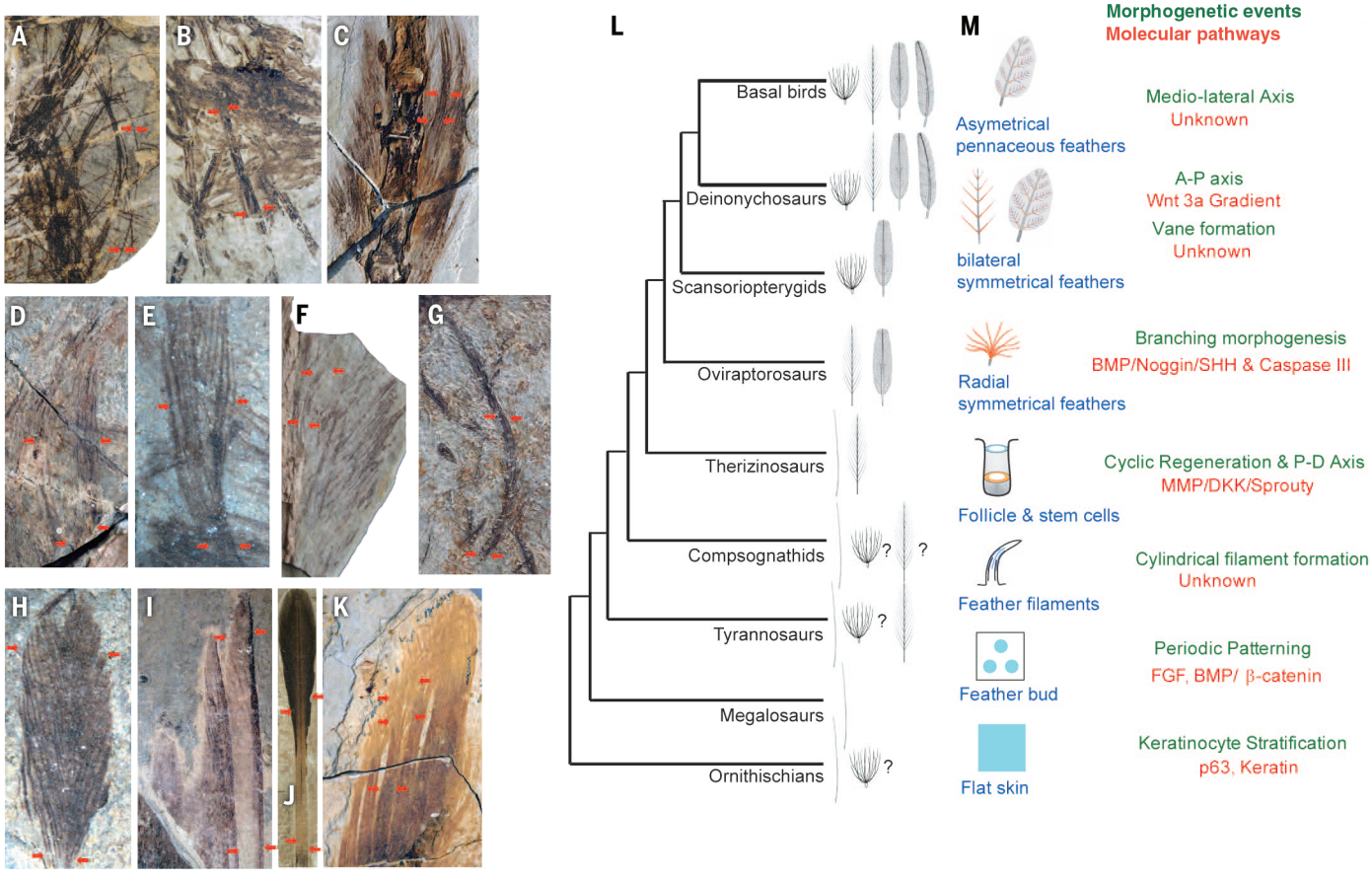


Fig. 5. The morphogenesis and evolution of feathers in dinosaurs. (**A**) Monofilamentous feathers in *Tianyulong*. (**B**) Broad monofilamentous feathers in *Beipiaosaurus*. Radially branched feathers in *Sinosauropteryx* (**C**), *Sinornithosaurus* (**D**), and *Anchiornis* (**E**). Bilaterally branched feathers in *Dilong* (**F**) and *Sinornithosaurus* (**G**). (**H**) Wing flight feathers with symmetrical vanes in *Anchiornis*. (**I**) Pedal flight feathers with asymmetrical vanes in *Microraptor*. (**J**) Rachis-dominant tail feathers in *Confuciusornis*. (**K**) Proximally ribbonlike tail feathers in *Similicaudipteryx*. (**L**) Phylogenetic distribution of major feather morphotypes (monofilamentous, radially branched, bilaterally branched, symmetrical flight, and asymmetrical flight feathers) among dinosaurs. (**M**) Major novel morphogenetic events and molecular pathways during feather evolution. These major feather morphotypes can be explained by selective usage of the five novel "molecular circuits" discussed in the text. Red arrows flank a feather.

subsequent caspase-dependent selective apoptosis can open up the tube, allowing feather vane formation (*128*).

(ii) Hierarchical branching and barb/barbule pattern formation. Ratio of local activators (noggin, Shh, and so on) and inhibitors (bone morphogenetic proteins) regulate the number and size of barb versus rachidial ridges (*5*, *129*). Barb ridges, which depend on Wnt 3a gradient, can run parallel to each other to form radial symmetric feathers or angle toward the rachis to form bilateral symmetric feathers (*130*).

(iii) Follicle formation and regenerative cycling. During development, matrix maetalloproteinase–dependent tissue invagination allows epithelia to topologically fold into a follicle (*131*). Follicle configuration allows protection of feather progenitors. Stem cells and dermal papilla are now clustered near the follicle base, and can undergo dickkopf Wnt–dependent cyclic regeneration (*132*).

(iv) Temporal regulation of stem cells along the proximal-distal axis. During feather formation, the molecular microenvironment and morphology can change along the distal (formed earlier) and proximal (formed later) through modulation of fibroblast growth factor–sprout activity (*133*).

(v) Organ metamorphosis and regional specific feather types. Upon regeneration, modulated by the body macroenvironments, different feather morphotypes (downy, contour, and flight feathers) can form on different body regions and in different physiological stages for optimal performance (*134*).

The aforementioned morphogenetic processes can be assembled selectively, in various orders, and independently to produce basic-to–more complex feather morphotypes (Fig. 5). The monofilamentous feathers seen in ornithischians (*54*, *56*, *57*) and some theropods (*55*, *57*), two sets of morphologically different flight feathers during ontogeny in the oviraptorosaurian *Similicaudipteryx* (*124*), and the extensive leg feathering (particularly pedal feathering) in a number of theropods and early birds (*135*) imply a variety of rearrangements in the morphogenetic modules. The wide occurrence of pedal feathering in theropods and early birds (*135*) and an extensive distribution of scales on the bodies of both feathered theropods (e.g., *Juravenator*) and feathered ornithischians (e.g., the ceratopsian *Psittacosaurus* and the ornithopod *Kullindadromeus*) remain particularly interesting from both developmental and evolutionary perspectives (*56*, *135*), and will no doubt play into further discussions on feather-scale relationships.

The diverse forms of theropod feathers apparently suggest a comparable myriad of functions. Initial functions of feathers more likely include thermoregulation and communication rather than flight (*57*). Several different lines of data highlight the importance of communication in early feather evolution (*41*, *57*, *117*, *136*), particularly by color patterns inferred via preserved melanosome morphologies in some theropods and early birds (*136*, *137*), although several studies have questioned the identifications of melanosomes in specimens of feathered theropods and early birds (*48*, *138*). One study shows an increased diversity of melanosome morphologies associated with the appearance of complex pinnate feathers during theropod evolution (*72*). This finding supports developmental data showing that unique topological arrangements allowing feathers to generate complex forms also permits the generation of visibly complex pigment patterns and pattern renewal during sexual maturation (*139*).

Four-winged dinosaurs and the origins of aerial behavior

Several flight-related anatomical features, such as hollow bones and the furcula, originated in early theropods; basal paravians had many hallmark features necessary for flight, including extremely small body size (*50*, *70*); a laterally oriented, long, and robust forelimb (*22*, *67*); an enlarged forebrain and other derived neurological adaptations (*71*); and large flight feathers (Figs. 1 and 2). Particularly surprising are the recent discoveries of large flight feathers forming a planar surface on the legs of some basal paravians—for example, those with asymmetrical vanes on both the tibia and metatarsus of some basal dromaeosaurs, such as *Microraptor* (*59*); large feathers with symmetrical vanes on both the tibia and metatarsus of the troodontid *Anchiornis* (*45*), the basal bird *Sapeornis*, and several other basal paravians (*135*); and large vaned feathers on tibiae of several basal birds including *Archaeopteryx*, confuciusornithids, and enantiornithines (*135*). These structures clearly would have been relevant to flight origins. The evolution of flight, nonetheless, remains highly debated, with a few recently proposed scenarios including factors as diverse as muscle hyperplasia (*140*) and the use of wings to assist in cursorial ascent of inclines, based on neontological observations (*11*). Multifactorial explanations for the origins of flight are likely, as this behavior has not been uniquely categorized; given variability in wing and tail configurations among theropods and early birds (*45*, *59*, *135*, *141*, *142*), flight probably had complex biomechanical origins (*142*, *143*).

Fundamental to all hypotheses of wing evolution, however, must be consideration of those circumstances that elicited aerial behavior in ancestrally terrestrial taxa. In all extant vertebrate and invertebrate gliders, flight is associated with gravitational acceleration downward from heights and with the use in aerodynamic control of diverse body structures, including but not limited to limbs, once aloft (*143*). The discoveries of four-winged theropods and early birds (Fig. 6) are consistent with the gravity-assisted hypothesis for the origin of bird flight. Moreover, the concurrent presence of a large and flattened tail, along with potential use of the feathered hind legs in aerodynamic control, suggests substantial capacity for using these as rudders and for maneuvering even during the earliest stages of bird evolution (*10*, *144*). Aerial righting reflexes, targeting movement while aloft, and controlled landings all would have been enhanced via bilaterally asymmetric motions of either wing pair. In other words, incipient and nonequilibrium flight behaviors would potentially associate with a four-winged and long-tailed arrangement in theropods and early birds (Fig. 6). There still remains disagreement as to the definitional criteria used to indicate the four-winged and long-tailed condition and as to whether large flight feathers with symmetrical

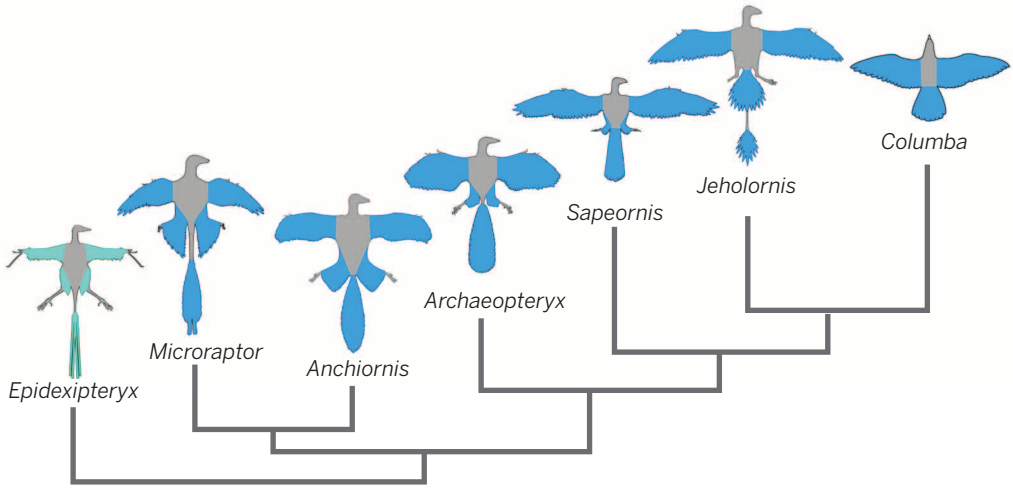


Fig. 6. Airfoils of selected maniraptorans. Airfoils (in blue) of the basal dromaeosaur *Microraptor*, the basal troodontid *Anchiornis*, the basal birds *Archaeopteryx*, *Sapeornis*, and *Jeholornis*, and a modern pigeon. Feathered airfoils are greater in number and are more distally positioned in early birds and their closest theropod relatives than in modern birds. The scansoriopterygid *Epidexipteryx* has stiff downlike feathers (in green) on both the forelimb and hindlimb and long ribbonlike feathers on the tail, the aerodynamic functions of which remain poorly known.

vanes in some basal paravians, including early birds, would create useful aerodynamic forces during aerial behaviors (*117*).

Use of the hind wings, and potentially the tail, to generate vertical force also results in a more posterior and stable position for the center of lift (*10*, *144*) (as can be deduced from bat and pterosaur wings based on patagial connections to the hindlegs). Such a hindwing configuration contributes to static stability in gliding, whereas subsequent functional shifts to the forewings alone, with associated reductions in tail size and leg feathering (*135*), would have required offsetting aerodynamic control via flapping motions (*145*). Potential tradeoffs between increased size of the forewings and the intensity of their kinematic activation (e.g., greater frequency and amplitude of flapping) are not well understood, although robotic models of early avialan morphologies can be informative (*146*). Such progressive anterior specialization of the flight apparatus, with enhanced forces produced by the emerging forelimb locomotor module (*145*), is nonetheless derived from an ancestrally four-winged platform (Fig. 6).

Future perspectives

Alternative phylogenetic hypotheses suggest a different evolutionary sequence of major bird characteristics and imply a different ecological origin of the group (*22*). Consequently a refined, more robust phylogeny will be imperative to move our studies forward. Increasing character and taxon sampling in recent theropod phylogenetic studies has, in theory, helped improve the accuracy of phylogenetic reconstructions, but better character formulation and more accurate scorings are imperative at present, pending better understanding of the morphologies of many recently discovered transitional forms.

In terms of character evolution, an integrative approach combining paleontological, neontological, developmental, geochronological, and even paleoenvironmental data is particularly desirable. A recent integrative study combining developmental, neontological, and paleontological data shows that birds evolved their highly specialized cranium mainly by paedomorphisis (*147*); another study using both neontological and paleontological data to reconstruct melanosome evolution suggests that feathers have changed significantly in their melanosome morphology among pennaraptoran theropods, which implies a physiological shift before the evolution of flight (*72*), and that concurs in association with extensive postcranial skeletal pneumaticity, which suggests the presence of birdlike endothermy (*29*). Similar approaches should be applied to other major structures. Certain avialan structures, such as feathers and visual system, are novel in both genetic and morphological terms. There are emerging data on how the genetic code affects the morphogenesis of these structures in developmental biology. How the genetic code and expression pattern evolved can be inferred by a detailed comparison of homologous structures in theropods, early birds and modern birds.

The construction of molecular phylogenies would enable a much better understanding of the genealogical relations among birds and their theropod relatives and the steps leading to the evolution of birds. Although once the topic of science-fiction movies, preserved organic remains have purportedly been recovered from *Tyrannosaurus* and the hadrosaur *Brachylophosaurus* (*148*, *149*) and from even older fossils such as Early Jurassic *Lufengosaurus* specimens (*150*). Greater examination along these lines in fossils spanning the theropod-bird transition is a potentially fruitful avenue for future investigation.

The increasing interest in the origins of birds from developmental biology and other disciplines has enriched our understanding of this important evolutionary event (*8*, *10*, *11*), and alternative evolutionary models on various aspects of the origin of birds have sometimes been provided based on neontological data (*4*, *11*). However, any evolutionary model must be tested using the fossil record. Consequently, dense fossil sampling along the line to birds and better understanding of the transitional forms play a key role in the discussion on this issue.

REFERENCES AND NOTES

1. A. Feduccia, *The Origin and Evolution of Birds* (Yale Univ. Press, New Haven, CT, ed. 2, 1999), pp. 466.
2. Z. Zhou, The origin and early evolution of birds: Discoveries, disputes, and perspectives from fossil evidence. *Naturwissenschaften* **91**, 455–471 (2004). doi: 10.1007/s00114-004-0570-4; pmid: 15365634
3. L. M. Witmer, in *Mesozoic Birds: Above the Heads of Dinosaurs*, L. M. Chiappe, L. M. Witmer, Eds. (Univ. of California Press, Berkeley, 2002), pp. 3–30.
4. R. O. Prum, Development and evolutionary origin of feathers. *J. Exp. Zool.* **285**, 291–306 (1999). doi: 10.1002/(SICI)1097-010X(19991215)285:4<291::AID-JEZ1>3.0.CO;2-9; pmid: 10578107
5. M. Yu, P. Wu, R. B. Widelitz, C.-M. Chuong, The morphogenesis of feathers. *Nature* **420**, 308–312 (2002). doi: 10.1038/nature01196; pmid: 12442169
6. A. O. Vargas, T. Kohlsdorf, J. F. Fallon, J. Vandenbrooks, G. P. Wagner, The evolution of *HoxD-11* expression in the bird wing: Insights from *Alligator mississippiensis*. *PLOS ONE* **3**, e3325 (2008). doi: 10.1371/journal.pone.0003325; pmid: 18833328
7. C. G. Farmer, K. Sanders, Unidirectional airflow in the lungs of alligators. *Science* **327**, 338–340 (2010). doi: 10.1126/science.1180219; pmid: 20075253
8. K. Tamura, N. Nomura, R. Seki, S. Yonei-Tamura, H. Yokoyama, Embryological evidence identifies wing digits in birds as digits 1, 2, and 3. *Science* **331**, 753–757 (2011). doi: 10.1126/science.1198229; pmid: 21311019
9. C. L. Organ, A. M. Shedlock, A. Meade, M. Pagel, S. V. Edwards, Origin of avian genome size and structure in non-avian dinosaurs. *Nature* **446**, 180–184 (2007). doi: 10.1038/nature05621; pmid: 17344851
10. D. Evangelista *et al.*, Aerodynamic characteristics of a feathered dinosaur measured using physical models. Effects of form on static stability and control effectiveness. *PLOS ONE* **9**, e85203 (2014). doi: 10.1371/journal.pone.0085203; pmid: 24454820
11. K. P. Dial, Wing-assisted incline running and the evolution of flight. *Science* **299**, 402–404 (2003). doi: 10.1126/science.1078237; pmid: 12532020
12. F. C. James, J. A. Pourtless IV, Cladistics and the origin of birds: A review and two new analyses. *Ornithol. Monogr.* **66**, 1–78 (2009). doi: 10.1525/om.2009.66.1.1
13. M. Norell, X. Xu, Feathered dinosaurs. *Annu. Rev. Earth Planet. Sci.* **33**, 277–299 (2005). doi: 10.1146/annurev.earth.33.092203.122511
14. J. Gauthier, Saurischian monophyly and the origin of birds. *Mem. Calif. Acad. Sci.* **8**, 1–55 (1986).
15. P. C. Sereno, The evolution of dinosaurs. *Science* **284**, 2137–2147 (1999). doi: 10.1126/science.284.5423.2137; pmid: 10381873
16. A. H. Turner, P. J. Makovicky, M. A. Norell, A review of dromaeosaurid systematics and paravian phylogeny. *Bull. Am. Mus. Nat. Hist.* **371**, 1–206 (2012). doi: 10.1206/748.1
17. P. Senter, J. I. Kirkland, D. D. DeBlieux, S. Madsen, N. Toth, New dromaeosaurids (Dinosauria: Theropoda) from the Lower Cretaceous of Utah, and the evolution of the dromaeosaurid tail. *PLOS ONE* **7**, e36790 (2012). doi: 10.1371/journal.pone.0036790; pmid: 22615813
18. T. R. Holtz Jr., A new phylogeny of the carnivorous dinosaurs. *Gaia* **15**, 5–61 (1998).
19. O. W. M. Rauhut, The interrelationships and evolution of basal theropod dinosaurs. *Spec. Pap. Paleontology* **69**, 1–213 (2003).
20. M. A. Norell, J. M. Clark, P. J. Makovicky, in *New Perspectives on the Origin and Evolution of Birds*, J. Gauthier, L. F. Gall, Eds. (Yale Univ. Press, New Haven, CT, 2001), pp. 49–67.
21. P. Godefroit *et al.*, A Jurassic avialan dinosaur from China resolves the early phylogenetic history of birds. *Nature* **498**, 359–362 (2013). doi: 10.1038/nature12168; pmid: 23719374
22. X. Xu, H. You, K. Du, F. Han, An *Archaeopteryx*-like theropod from China and the origin of Avialae. *Nature* **475**, 465–470 (2011). doi: 10.1038/nature10288; pmid: 21796204
23. J. N. Choiniere *et al.*, A basal alvarezsauroid theropod from the early Late Jurassic of Xinjiang, China. *Science* **327**, 571–574 (2010). doi: 10.1126/science.1182143; pmid: 20110503
24. J. H. Ostrom, The osteology of *Deinonychus antirrhopus*, an unusual theropod from the Lower Cretaceous of Montana. *Bull. Peabody Mus. Nat. Hist.* **30**, 1–165 (1969).
25. R. Barsbold, Predatory dinosaurs from the Cretaceous of Mongolia. *Tr. Sovm. Sov.-Mong. Paleontol. Eksped.* **19**, 1–117 (1983).
26. O. Wings, thesis, The University of Bonn (2004).
27. P. M. O'Connor, L. P. A. M. Claessens, Basic avian pulmonary design and flow-through ventilation in non-avian theropod dinosaurs. *Nature* **436**, 253–256 (2005). doi: 10.1038/nature03716; pmid: 16015329
28. P. C. Sereno *et al.*, Evidence for avian intrathoracic air sacs in a new predatory dinosaur from Argentina. *PLOS ONE* **3**, e3303 (2008). doi: 10.1371/journal.pone.0003303; pmid: 18825273
29. R. B. J. Benson, R. J. Butler, M. T. Carrano, P. M. O'Connor, Air-filled postcranial bones in theropod dinosaurs: Physiological implications and the 'reptile'-bird transition. *Biol. Rev. Camb. Philos. Soc.* **87**, 168–193 (2012). doi: 10.1111/j.1469-185X.2011.00190.x; pmid: 21733078
30. M. A. Norell, J. M. Clark, L. M. Chiappe, D. Dashzeveg, A nesting dinosaur. *Nature* **378**, 774–776 (1995). doi: 10.1038/378774a0
31. D. J. Varricchio, F. Jackson, J. J. Borkowski, J. R. Horner, Nest and egg clutches of the dinosaur *Troodon formosus* and the evolution of avian reproductive traits. *Nature* **385**, 247–250 (1997). doi: 10.1038/385247a0
32. G. M. Erickson, K. Curry Rogers, D. J. Varricchio, M. A. Norell, X. Xu, Growth patterns in brooding dinosaurs reveals the timing of sexual maturity in non-avian dinosaurs and genesis of the avian condition. *Biol. Lett.* **3**, 558–561 (2007). pmid: 17638674
33. G. Grellet-Tinner, P. Makovicky, A possible egg of the dromaeosaur *Deinonychus antirrhopus*: Phylogenetic and biological implications. *Can. J. Earth Sci.* **43**, 705–719 (2006). doi: 10.1139/e06-033
34. G. Grellet-Tinner, L. M. Chiappe, in *Feathered Dragons: Studies on the Transition from Dinosaurs to Birds*, P. J. Currie, E. B. Koppelhus, M. A. Shugar, J. L. Wright, Eds. (Indiana Univ. Press, Bloomington, 2004), pp. 185–214.
35. D. J. Varricchio, F. D. Jackson, A phylogenetic assessment of prismatic dinosaur eggs from the Cretaceous Two Medicine Formation of Montana. *J. Vertebr. Paleontol.* **24**, 931–937 (2004). doi: 10.1671/0272-4634(2004)024[0931:APAOPD]2.0.CO;2
36. D. K. Zelenitsky, F. Therrien, Phylogenetic analysis of reproductive traits of maniraptoran theropods and its implications for egg parataxonomy. *Paleontology* **51**, 807–816 (2008). doi: 10.1111/j.1475-4983.2008.00770.x
37. K. Padian, A. J. de Ricqlès, J. R. Horner, Dinosaurian growth rates and bird origins. *Nature* **412**, 405–408 (2001). doi: 10.1038/35086500; pmid: 11473307
38. F. L. Agnolín, F. E. Novas, *Avian Ancestors: A Review of the Phylogenetic Relationships of the Theropods Unenlagiidae, Microraptoria, Anchiornis, and Scansoriopterygidae* (Springer, Dordrecht, Heidelberg, 2013).
39. P. J. Makovicky, L. E. Zanno, in *Living Dinosaurs: The Evolutionary History of Modern Birds*, G. J. Dyke, G. Kaiser, Eds. (Wiley-Blackwell, Oxford, 2011), pp. 9–29.

40. C. A. Forster, S. D. Sampson, L. M. Chiappe, D. W. Krause, The theropod ancestry of birds: New evidence from the Late Cretaceous of Madagascar. *Science* **279**, 1915–1919 (1998). doi: 10.1126/science.279.5358.1915; pmid: 9506938
41. F. Zhang, Z. Zhou, X. Xu, X. Wang, C. Sullivan, A bizarre Jurassic maniraptoran from China with elongate ribbon-like feathers. *Nature* **455**, 1105–1108 (2008). doi: 10.1038/nature07447; pmid: 18948955
42. X. Xu *et al*., A new feathered maniraptoran dinosaur fossil that fills a morphological gap in avian origin. *Chin. Sci. Bull.* **54**, 430–435 (2009). doi: 10.1007/s11434-009-0009-6
43. F. E. Novas, P. F. Puertat, New evidence concerning avian origins from the Late Cretaceous of Patagonia. *Nature* **387**, 390–392 (1997). doi: 10.1038/387390a0
44. P. J. Makovicky, S. Apesteguía, F. L. Agnolín, The earliest dromaeosaurid theropod from South America. *Nature* **437**, 1007–1011 (2005). doi: 10.1038/nature03996; pmid: 16222297
45. D. Hu, L. Hou, L. Zhang, X. Xu, A pre-*Archaeopteryx* troodontid theropod from China with long feathers on the metatarsus. *Nature* **461**, 640–643 (2009). doi: 10.1038/nature08322; pmid: 19794491
46. X. Xu, Q.-Y. Ma, D.-Y. Hu, Pre-*Archaeopteryx* coelurosaurian dinosaurs and their implications for understanding avian origins. *Chin. Sci. Bull.* **55**, 3971–3977 (2010). doi: 10.1007/s11434-010-4150-z
47. J. O'Connor, C. Sullivan, *Vertebr. PalAsiat.* **52**, 3 (2014).
48. A. Feduccia, Bird origins anew. *Auk* **130**, 1–12 (2013). doi: 10.1525/auk.2013.130.1.1
49. J. A. Ruben *et al*., Pulmonary function and metabolic physiology of theropod dinosaurs. *Science* **283**, 514–516 (1999). doi: 10.1126/science.283.5401.514; pmid: 9915693
50. A. H. Turner, D. Pol, J. A. Clarke, G. M. Erickson, M. A. Norell, A basal dromaeosaurid and size evolution preceding avian flight. *Science* **317**, 1378–1381 (2007).pmid: 17823350
51. X. Xu, M. A. Norell, X. L. Wang, P. J. Makovicky, X. C. Wu, A basal troodontid from the Early Cretaceous of China. *Nature* **415**, 780–784 (2002). doi: 10.1038/415780a; pmid: 11845206
52. X. Xu *et al*., A Jurassic ceratosaur from China helps clarify avian digital homologies. *Nature* **459**, 940–944 (2009). doi: 10.1038/nature08124; pmid: 19536256
53. P. Chen, Z. Dong, S. Zhen, An exceptionally well-preserved theropod dinosaur from the Yixian Formation of China. *Nature* **391**, 147–152 (1998). doi: 10.1038/34356
54. X. T. Zheng, H. L. You, X. Xu, Z. M. Dong, An Early Cretaceous heterodontosaurid dinosaur with filamentous integumentary structures. *Nature* **458**, 333–336 (2009). doi: 10.1038/nature07856; pmid: 19295609
55. O. W. M. Rauhut, C. Foth, H. Tischlinger, M. A. Norell, Exceptionally preserved juvenile megalosauroid theropod dinosaur with filamentous integument from the Late Jurassic of Germany. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 11746–11751 (2012). doi: 10.1073/pnas.1203238109; pmid: 22753486
56. P. Godefroit *et al*., A Jurassic ornithischian dinosaur from Siberia with both feathers and scales. *Science* **345**, 451–455 (2014). doi: 10.1126/science.1253351; pmid: 25061209
57. X. Xu, Y. Guo, *Vertebr. PalAsiat.* **47**, 311 (2009).
58. Q. Ji, P. J. Currie, M. A. Norell, S. A. Ji, Two feathered dinosaurs from northeastern China. *Nature* **393**, 753 (1998). doi: 10.1038/31635
59. X. Xu *et al*., Four-winged dinosaurs from China. *Nature* **421**, 335–340 (2003). doi: 10.1038/nature01342; pmid: 12540892
60. X. Xu, M. A. Norell, A new troodontid dinosaur from China with avian-like sleeping posture. *Nature* **431**, 838–841 (2004). doi: 10.1038/nature02898; pmid: 15483610
61. G. Mayr, B. Pohl, S. Hartman, D. S. Peters, The tenth skeletal specimen of *Archaeopteryx*. *Zool. J. Linn. Soc.* **149**, 97–116 (2007). doi: 10.1111/j.1096-3642.2006.00245.x
62. Z. Zhou, F. Zhang, A long-tailed, seed-eating bird from the Early Cretaceous of China. *Nature* **418**, 405–409 (2002). doi: 10.1038/nature00930; pmid: 12140555
63. Z. H. Zhou, F. C. Zhang, Anatomy of the primitive bird *Sapeornis chaoyangensis* from the Early Cretaceous of Liaoning, China. *Can. J. Earth Sci.* **40**, 731–747 (2003). doi: 10.1139/e03-011
64. C. A. Brochu, M. A. Norell, in *New Perspectives on the Origin and Early Evolution of Birds*, J. A. Gauthier, L. F. Gall, Eds. (Peabody Museum of Natural History, Yale Univ., New Haven, CT, 2001), pp. 511–536.
65. X. Xu *et al*., A basal tyrannosauroid dinosaur from the Late Jurassic of China. *Nature* **439**, 715–718 (2006). doi: 10.1038/nature04511; pmid: 16467836
66. J. M. Clark, T. Maryanska, R. Barsbold, in *The Dinosauria*, D. B. Weishampel, P. Dodson, H. Osmolska, Eds. (Univ. of California Press, Berkeley, ed. 2, 2004), pp. 151–164.
67. V. Allen, K. T. Bates, Z. Li, J. R. Hutchinson, Linking the evolution of body shape and locomotor biomechanics in bird-line archosaurs. *Nature* **497**, 104–107 (2013). doi: 10.1038/nature12059; pmid: 23615616
68. C. Lipkin, P. C. Sereno, J. R. Horner, The furcula in *Suchomimus tenerensis* and *Tyrannosaurus rex* (Dinosauria: Theropoda: Tetanurae). *J. Paleontol.* **81**, 1523–1527 (2007). doi: 10.1666/06-024.1
69. X. Xu, F. Han, Q. Zhao, Homologies and homeotic transformation of the theropod 'semilunate' carpal. *Sci. Rep.* **4**, 6042 (2014).pmid: 25116378
70. M. S. Y. Lee, A. Cau, D. Naish, G. J. Dyke, Sustained miniaturization and anatomical innovation in the dinosaurian ancestors of birds. *Science* **345**, 562–566 (2014). doi: 10.1126/science.1252243; pmid: 25082702
71. A. M. Balanoff, G. S. Bever, T. B. Rowe, M. A. Norell, Evolutionary origins of the avian brain. *Nature* **501**, 93–96 (2013). doi: 10.1038/nature12424; pmid: 23903660
72. Q. Li *et al*., Melanosome evolution indicates a key physiological shift within feathered dinosaurs. *Nature* **507**, 350–353 (2014). doi: 10.1038/nature12973; pmid: 24522537
73. K. E. Mikhailov, Fossil and recent eggshell in amniotic vertebrates: fine structure, comparative morphology and classification. *Spec. Pap. Palaeontol.* **56**, 1–80 (1997).
74. D. J. Varricchio, D. E. Barta, Revisiting Sabath's "larger avian eggs" from the Gobi Cretaceous. *Acta Palaeontol. Pol.* 10.4202/app.00085.2014 (2014).
75. T. Sato, Y.-N. Cheng, X.-C. Wu, D. K. Zelenitsky, Y.-F. Hsiao, A pair of shelled eggs inside a female dinosaur. *Science* **308**, 375 (2005). doi: 10.1126/science.1110578; pmid: 15831749
76. T. He, D. J. Varricchio, F. Jackson, X. Jin, A. W. Poust, *J. Vertebr. Paleontol.* **3,** (abstr.) 108 (2012).
77. F. Fanti, P. J. Currie, D. Badamgarav, New specimens of Nemegtomaia from the Baruungoyot and Nemegt formations (Late Cretaceous) of Mongolia. *PLOS ONE* **7**, e31330 (2012). doi: 10.1371/journal.pone.0031330; pmid: 22347465
78. D. J. Varricchio, F. D. Jackson, Origins of avian reproduction: Answers and questions from dinosaurs. *Palaeovertebrata* **32**, 149 (2003).
79. G. F. Birchard, M. Ruta, D. C. Deeming, Evolution of parental incubation behaviour in dinosaurs cannot be inferred from clutch mass in birds. *Biol. Lett.* **9**, 20130036 (2013). doi: 10.1098/rsbl.2013.0036
80. S. L. Vehrencamp, Evolutionary routes to joint-female nesting in birds. *Behav. Ecol.* **11**, 334–344 (2000). doi: 10.1093/beheco/11.3.334
81. D. J. Varricchio *et al*., Avian paternal care had dinosaur origin. *Science* **322**, 1826–1828 (2008). doi: 10.1126/science.1163245; pmid: 19095938
82. F. L. Agnolin, J. E. Powell, F. E. Novas, M. Kundrát, New alvarezsaurid (Dinosauria, Theropoda) from uppermost Cretaceous of north-western Patagonia with associated eggs. *Cretac. Res.* **35**, 33–56 (2012). doi: 10.1016/j.cretres.2011.11.014
83. D. J. Varricchio, J. R. Horner, F. D. Jackson, Embryos and eggs for the Cretaceous theropod dinosaur *Troodon formosus*. *J. Vertebr. Paleontol.* **22**, 564–576 (2002). doi: 10.1671/0272-4634(2002)022[0564:EAEFTC]2.0.CO;2
84. F. D. Jackson, J. R. Horner, D. J. Varricchio, A study of a *Troodon* egg containing embryonic remains using epifluorescence microscopy and other techniques. *Cretac. Res.* **31**, 255–262 (2010). doi: 10.1016/j.cretres.2009.11.006
85. D. J. Varricchio, F. Jackson, C. N. Trueman, A nesting trace with eggs for the Cretaceous theropod dinosaur *Troodon formosus*. *J. Vertebr. Paleontol.* **19**, 91–100 (1999). doi: 10.1080/02724634.1999.10011125
86. D. J. Varricchio, F. D. Jackson, R. A. Jackson, D. K. Zelenitsky, Porosity and water vapor conductance of two *Troodon formosus* eggs: An assessment of incubation strategy in a maniraptoran dinosaur. *Paleobiology* **39**, 278–296 (2013). doi: 10.1666/11042
87. K. Sabath, *Acta Palaeontol. Pol.* **36**, 151 (1991).
88. N. López-Martínez, E. Vicens, A new peculiar dinosaur egg, *Sankofa pyrenaica* oogen. nov. oosp. nov. from the Upper Cretaceous coastal deposits of the Aren Formation, south-central Pyrenees, Lleida, Catalonia, Spain. *Paleontology* **55**, 325–339 (2012). doi: 10.1111/j.1475-4983.2011.01114.x
89. G. Mayr, A. Manegold, Can ovarian follicles fossilize? *Nature* **499**, E1 (2013). doi: 10.1038/nature12367; pmid: 23846661
90. X. Zheng *et al*., Preservation of ovarian follicles reveals early evolution of avian reproductive behaviour. *Nature* **495**, 507–511 (2013). doi: 10.1038/nature11985; pmid: 23503663
91. T. P. Hopp, M. J. Orsen, in *Feathered Dragons: Studies on the Transition from Dinosaurs to Birds*, P. J. Currie, E. B. Koppelhus, M. A. Shugar, J. L. Wright, Eds. (Indiana Univ. Press, 2004), pp. 234–250.
92. D. J. Varricchio, F. D. Jackson, in *Feathered Dragons: Studies on the Transition from Dinosaurs to Birds*, P. J. Currie, E. B. Koppelhus, M. A. Shugar, J. L. Wright, Eds. (Indiana Univ. Press, Bloomington, 2004), pp. 215–233.
93. J. K. Terres, *The Audubon Society Encyclopedia of North American Birds* (Wings Books, New York, 1995).
94. D. C. Deeming, in *Egg Incubation*, D. C. Deeming, M. W. J. Ferguson, Eds. (Cambridge Univ. Press, Cambridge, 1991), pp. 307–323.
95. I. W. A. Calder, *Size, Function, and Life History* (Harvard Univ. Press, Cambridge, 1984), pp. 431.
96. A. Chinsamy, L. Chiappe, P. Dodson, Growth rings in Mesozoic birds. *Nature* **368**, 196–197 (1994). doi: 10.1038/368196a0
97. G. M. Erickson, K. C. Rogers, S. A. Yerby, Dinosaurian growth patterns and rapid avian growth rates. *Nature* **412**, 429–433 (2001). doi: 10.1038/35086558; pmid: 11473315
98. G. M. Erickson *et al*., Was dinosaurian physiology inherited by birds? Reconciling slow growth in *Archaeopteryx*. *PLOS ONE* **4**, e7390 (2009). doi: 10.1371/journal.pone.0007390; pmid: 19816582
99. A. Chinsamy, L. M. Chiappe, J. Marugán-Lobón, G. Chunling, Z. Fengjiao, Gender identification of the Mesozoic bird *Confuciusornis sanctus*. *Nat. Commun.* **4**, 1381 (2013). doi: 10.1038/ncomms2377; pmid: 23340421
100. A. Chinsamy, in *Mesozoic Birds: Above the Heads of Dinosaurs*, L. M. Chiappe, L. M. Witmer, Eds. (Univ. of California Press, Berkeley, 2002), pp. 421–431.
101. N. P. Myhrvold, Revisiting the estimation of dinosaur growth rates. *PLOS ONE* **8**, e81917 (2013). doi: 10.1371/journal.pone.0081917; pmid: 24358133
102. J. M. Grady, B. J. Enquist, E. Dettweiler-Robinson, N. A. Wright, F. A. Smith, Evidence for mesothermy in dinosaurs. *Science* **344**, 1268–1272 (2014). doi: 10.1126/science.1253143; pmid: 24926017
103. G. P. Wagner, J. A. Gauthier, 1,2,3 = 2,3,4: A solution to the problem of the homology of the digits in the avian hand. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 5111–5116 (1999). doi: 10.1073/pnas.96.9.5111; pmid: 10220427
104. M. A. G. de Bakker *et al*., Digit loss in archosaur evolution and the interplay between selection and constraints. *Nature* **500**, 445–448 (2013). doi: 10.1038/nature12336; pmid: 23831646
105. A. C. Burke, A. Feduccia, Developmental patterns and the identification of homologies in the avian hand. *Science* **278**, 666–668 (1997). doi: 10.1126/science.278.5338.666
106. Z. Wang, R. L. Young, H. Xue, G. P. Wagner, Transcriptomic analysis of avian digits reveals conserved and derived digit identities in birds. *Nature* **477**, 583–586 (2011). doi: 10.1038/nature10391; pmid: 21892187
107. M. Towers, J. Signolet, A. Sherman, H. Sang, C. Tickle, Insights into bird wing evolution and digit specification from polarizing region fate maps. *Nat. Commun.* **2**, 426 (2011). doi: 10.1038/ncomms1437; pmid: 21829188
108. S. Chatterjee, Counting the fingers of birds and dinosaurs. *Science* **280**, 355 (1998). doi: 10.1126/science.280.5362.355a
109. N. H. Shubin, P. Alberch, *Evol. Biol.* **20**, 319 (1986).
110. X. Xu, S. Mackem, Tracing the evolution of avian wing digits. *Curr. Biol.* **23**, R538 (2013). doi: 10.1016/j.cub.2013.04.071
111. P. J. Scherz, E. McGlinn, S. Nissim, C. J. Tabin, Extended exposure to Sonic hedgehog is required for patterning the posterior digits of the vertebrate limb. *Dev. Biol.* **308**, 343–354 (2007). doi: 10.1016/j.ydbio.2007.05.030; pmid: 17610861
112. J. Zhu *et al*., Uncoupling Sonic hedgehog control of pattern and expansion of the developing limb bud. *Dev. Cell* **14**, 624–632 (2008). doi: 10.1016/j.devcel.2008.01.008; pmid: 18410737
113. E. R. Schachner, J. R. Hutchinson, C. Farmer, Pulmonary anatomy in the Nile crocodile and the evolution of unidirectional airflow in Archosauria. *PeerJ.* **1**, e60 (2013). doi: 10.7717/peerj.60; pmid: 23638399
114. E. R. Schachner, R. L. Cieri, J. P. Butler, C. G. Farmer, Unidirectional pulmonary airflow patterns in the savannah monitor lizard. *Nature* **506**, 367–370 (2014). doi: 10.1038/nature12871; pmid: 24336209
115. A. H. Turner, P. J. Makovicky, M. A. Norell, Feather quill knobs in the dinosaur *Velociraptor*. *Science* **317**, 1721 (2007). doi: 10.1126/science.1145076; pmid: 17885130
116. F. Ortega, F. Escaso, J. L. Sanz, A bizarre, humped Carcharodontosauria (Theropoda) from the lower cretaceous of Spain. *Nature* **467**, 203–206 (2010). doi: 10.1038/nature09181; pmid: 20829793

117. C. Foth, H. Tischlinger, O. W. M. Rauhut, New specimen of *Archaeopteryx* provides insights into the evolution of pennaceous feathers. *Nature* **511**, 79–82 (2014). doi: 10.1038/nature13467; pmid: 24990749
118. R. Barsbold, H. Osmolska, M. Watabe, P. J. Currie, K. Tsogtbaatar, A new oviraptorosaur (Dinosauria, Theropoda) from Mongolia: The first dinosaur with a pygostyle. *Acta Palaeontol. Pol.* **45**, 97–106 (2000).
119. M. H. Schweitzer *et al*., Beta-keratin specific immunological reactivity in feather-like structures of the cretaceous alvarezsaurid, *Shuvuuia deserti*. *J. Exp. Zool.* **285**, 146–157 (1999). doi: 10.1002/(SICI)1097-010X(19990815)285:2<146::AID-JEZ7>3.0.CO;2-A; pmid: 10440726
120. Q. Ji, S. A. Ji, On the discovery of the earliest fossil bird in China (*Sinosauropteryx* gen. nov.) and the origin of birds. *Chinese Geol.* **233**, 30–33 (1996).
121. U. B. Göhlich, L. M. Chiappe, A new carnivorous dinosaur from the Late Jurassic Solnhofen archipelago. *Nature* **440**, 329–332 (2006). doi: 10.1038/nature04579; pmid: 16541071
122. D. K. Zelenitsky *et al*., Feathered non-avian dinosaurs from North America provide insight into wing origins. *Science* **338**, 510–514 (2012). doi: 10.1126/science.1225376; pmid: 23112330
123. R. C. McKellar, B. D. E. Chatterton, A. P. Wolfe, P. J. Currie, A diverse assemblage of Late Cretaceous dinosaur and bird feathers from Canadian amber. *Science* **333**, 1619–1622 (2011). doi: 10.1126/science.1203344; pmid: 21921196
124. X. Xu, X. Zheng, H. You, Exceptional dinosaur fossils show ontogenetic development of early feathers. *Nature* **464**, 1338–1341 (2010). doi: 10.1038/nature08965; pmid: 20428169
125. C.F. Chen *et al*., Development, regeneration and evolution of feathers. *Annu. Rev. Anim. Biosci.* (2014). pmid: 25387232
126. Z. Yue, T. X. Jiang, R. B. Widelitz, C. M. Chuong, Mapping stem cell activities in the feather follicle. *Nature* **438**, 1026–1029 (2005). doi: 10.1038/nature04222; pmid: 16355227
127. R. Chodankar *et al*., Shift of localized growth zones contributes to skin appendage morphogenesis: Role of the Wnt/β-catenin pathway. *J. Invest. Dermatol.* **120**, 20–26 (2003). doi: 10.1046/j.1523-1747.2003.12008.x
128. C. H. Chang *et al*., Sculpting skin appendages out of epidermal layers via temporally and spatially regulated apoptotic events. *J. Invest. Dermatol.* **122**, 1348–1355 (2004). doi: 10.1111/j.0022-202X.2004.22611.x; pmid: 15175023
129. M. P. Harris, J. F. Fallon, R. O. Prum, Shh-Bmp2 signaling module and the evolutionary origin and diversification of feathers. *J. Exp. Zool.* **294**, 160–176 (2002). doi: 10.1002/jez.10157; pmid: 12210117
130. Z. Yue, T. X. Jiang, R. B. Widelitz, C. M. Chuong, Wnt3a gradient converts radial to bilateral feather symmetry via topological arrangement of epithelia. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 951–955 (2006). doi: 10.1073/pnas.0506894103; pmid: 16418297
131. T. X. Jiang, T. L. Tuan, P. Wu, R. B. Widelitz, C. M. Chuong, From buds to follicles: Matrix metalloproteinases in developmental tissue remodeling during feather morphogenesis. *Differentiation* **81**, 307–314 (2011). doi: 10.1016/j.diff.2011.03.004; pmid: 21497985
132. Q. Chu *et al*., Dkk2/Frzb in the dermal papillae regulates feather regeneration. *Dev. Biol.* **387**, 167–178 (2014). doi: 10.1016/j.ydbio.2014.01.010; pmid: 24463139
133. Z. Yue, T. X. Jiang, P. Wu, R. B. Widelitz, C. M. Chuong, Sprouty/FGF signaling regulates the proximal-distal feather morphology and the size of dermal papillae. *Dev. Biol.* **372**, 45–54 (2012). doi: 10.1016/j.ydbio.2012.09.004; pmid: 23000358
134. C. M. Chuong, V. A. Randall, R. B. Widelitz, P. Wu, T. X. Jiang, Physiological regeneration of skin appendages and implications for regenerative medicine. *Physiology (Bethesda)* **27**, 61–72 (2012). doi: 10.1152/physiol.00028.2011; pmid: 22505663
135. X. Zheng *et al*., Hind wings in basal birds and the evolution of leg feathers. *Science* **339**, 1309–1312 (2013). doi: 10.1126/science.1228753; pmid: 23493711
136. Q. Li *et al*., Reconstruction of *Microraptor* and the evolution of iridescent plumage. *Science* **335**, 1215–1219 (2012). doi: 10.1126/science.1213780; pmid: 22403389
137. F. Zhang *et al*., Fossilized melanosomes and the colour of Cretaceous dinosaurs and birds. *Nature* **463**, 1075–1078 (2010). doi: 10.1038/nature08740; pmid: 20107440
138. A. E. Moyer *et al*., Melanosomes or microbes: Testing an alternative hypothesis for the origin of microbodies in fossil feathers. *Sci. Rep.* **4**, 4233 (2014). doi: 10.1038/srep04233; pmid: 24595214
139. S. J. Lin *et al*., Topology of feather melanocyte progenitor niche allows complex pigment patterns to emerge. *Science* **340**, 1442–1445 (2013). doi: 10.1126/science.1230374; pmid: 23618762
140. S. A. Newman, Thermogenesis, muscle hyperplasia, and the origin of birds. *BioEssays* **33**, 653–656 (2011). doi: 10.1002/bies.201100061; pmid: 21695679
141. J. O'Connor *et al*., Unique caudal plumage of *Jeholornis* and complex tail evolution in early birds. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 17404–17408 (2013). doi: 10.1073/pnas.1316979110; pmid: 24101506
142. N. R. Longrich, J. Vinther, Q. Meng, Q. Li, A. P. Russell, Primitive wing feather arrangement in *Archaeopteryx lithographica* and *Anchiornis huxleyi*. *Curr. Biol.* **22**, 2262–2267 (2012). doi: 10.1016/j.cub.2012.09.052; pmid: 23177480
143. R. Dudley *et al*., Gliding and the functional origins of flight: Biomechanical novelty or necessity? *Annu. Rev. Ecol. Evol. Syst.* **38**, 179–201 (2007). doi: 10.1146/annurev.ecolsys.37.091305.110014
144. G. Dyke *et al*., Aerodynamic performance of the feathered dinosaur *Microraptor* and the evolution of feathered flight. *Nat. Commun.* **4**, 2489 (2013). doi: 10.1038/ncomms3489; pmid: 24048346
145. S. M. Gatesy, K. P. Dial, Locomotor modules and the evolution of avian flight. *Evolution* **50**, 331 (1996). doi: 10.2307/2410804
146. K. Peterson, P. Birkmeyer, R. Dudley, R. S. Fearing, A wing-assisted running robot and implications for avian flight evolution. *Bioinspir. Biomim.* **6**, 046008 (2011). doi: 10.1088/1748-3182/6/4/046008; pmid: 22004831
147. B. A. S. Bhullar *et al*., Birds have paedomorphic dinosaur skulls. *Nature* **487**, 223–226 (2012). doi: 10.1038/nature11146; pmid: 22722850
148. J. M. Asara, M. H. Schweitzer, L. M. Freimark, M. Phillips, L. C. Cantley, Protein sequences from mastodon and *Tyrannosaurus rex* revealed by mass spectrometry. *Science* **316**, 280–285 (2007). doi: 10.1126/science.1137614; pmid: 17431180
149. M. H. Schweitzer *et al*., Biomolecular characterization and protein sequences of the Campanian hadrosaur *B. canadensis*. *Science* **324**, 626–631 (2009). doi: 10.1126/science.1165069; pmid: 19407199
150. R. R. Reisz *et al*., Embryology of Early Jurassic dinosaur from China with evidence of preserved organic remains. *Nature* **496**, 210–214 (2013). doi: 10.1038/nature11978; pmid: 23579680

ACKNOWLEDGMENTS

We thank Y. Liu, L. Xing, R. Li, X. Ding, P. M. O'Connor, M. Ellison, H. Zang, and K. Womble for illustrations. X.X. and Z.H.Z. thank X. Zheng (Shandong Tianyu Museum of Nature) and H. Li (Jizantang Museum) for access to specimens in their care. D.J.V. thanks the Fukui Prefecture Dinosaur Museum for research support. X.X. and Z.H.Z. are supported by the National Natural Science Foundation of China (41120124002) and 973 program (2012CB821900), C.M.C. by U.S. National Institute of Arthritis and Musculoskeletal Diseases, NIH, grant AR 60306 and 47364, G.M.E. by NSF grants DBI 0446224 and EAR 044186549, and D.J.V. by NSF grant EAR 0847777.

10.1126/science.1253293

REPORTS

ROCK MECHANICS

Superplastic nanofibrous slip zones control seismogenic fault friction

Berend A. Verberne,[1]* Oliver Plümper,[1] D.A. Matthijs de Winter,[2] Christopher J. Spiers[1]

Understanding the internal mechanisms controlling fault friction is crucial for understanding seismogenic slip on active faults. Displacement in such fault zones is frequently localized on highly reflective (mirrorlike) slip surfaces, coated with thin films of nanogranular fault rock. We show that mirror-slip surfaces developed in experimentally simulated calcite faults consist of aligned nanogranular chains or fibers that are ductile at room conditions. These microstructures and associated frictional data suggest a fault-slip mechanism resembling classical Ashby-Verrall superplasticity, capable of producing unstable fault slip. Diffusive mass transfer in nanocrystalline calcite gouge is shown to be fast enough for this mechanism to control seismogenesis in limestone terrains. With nanogranular fault surfaces becoming increasingly recognized in crustal faults, the proposed mechanism may be generally relevant to crustal seismogenesis.

The cores of seismically active fault zones are frequently characterized by a narrow (millimeters to centimeters thick) principal slip zone (PSZ) composed of nanogranular fault rock (*1*, *2*). Recent field studies demonstrate that fault-parallel exposures of such PSZs often form highly reflective (specular) surfaces, suggesting that these "mirrorlike" fault surfaces may be indicators of past seismic slip (*3*–*7*). Independently of fault plane morphology, for earthquake ruptures to nucleate, the internal slip zone wear product, or "fault gouge," must show a decrease in frictional strength with increasing fault slip rate. In other words, it must show runaway velocity- (or v-) weakening behavior, as opposed to self-stabilizing, v-strengthening slip (*8*). Elucidating the mechanisms controlling v-weakening and v-strengthening slip, especially those involving the formation and slip behavior of mirrorlike nanogranular fault surfaces, is therefore crucial for understanding the processes leading to earthquake nucleation.

Low-velocity ($v \approx 10^{-6}$ m/s) friction experiments, designed to investigate the onset of seismogenesis in the upper crust, recently demonstrated a transition from v-strengthening to v-weakening slip in simulated calcite fault gouge at temperatures above 80° to 100°C (*9*–*11*). Microstructural studies of the sheared samples revealed the development of continuous 10- to 100-μm-wide nanocrystalline PSZs with a crystallographic preferred orientation (CPO) (*10*). When split along the shear plane, these zones become partially exposed to display multiple patches of striated, highly reflective slip surfaces (*11*). Similar though more continuous nanogranular mirror-slip surfaces have been reported to form in high-velocity friction experiments achieving coseismic slip rates (v = 0.1 to 3 m/s) (*4*, *5*, *12*). At these slip rates, nanograins are believed to facilitate powder lubrication, causing strong dynamic fault weakening (*13*, *14*), but during slow, pre-seismic and interseismic slip their role remains enigmatic. Nanophase materials are well known to exhibit much faster diffusion and mass transport than their coarser-grained counterparts (*15*, *16*). Moreover, a recently proposed microphysical model for (thermally activated) v-weakening shear of fault gouge undergoing granular flow demonstrates that the rate of interparticle diffusive mass transfer can play a central role in controlling gouge strength and its velocity dependence (supplementary text and fig. S1) (*17*). Therefore, the nanophysical deformation mechanisms that operate in shiny PSZs may hold the key to understanding the processes controlling the initiation of seismic slip.

We investigated the micro- and nanostructural properties of PSZs developed in simulated calcite fault gouge subjected to direct shear at low slip rates (fig. S2 and table S1). The experiments were conducted on nominally dry samples at v = 1 μm/s at a normal stress of 50 MPa, at temperatures of 18° and 140°C (*18*). These temperatures respectively correspond to non-seismogenic, v-strengthening conditions and seismogenic, v-weakening conditions (*9*–*11*). Upon disassembly after each experiment, PSZs formed near the sheared sample boundary split open to reveal multiple shiny patches within the ultrafine-grained PSZ (Fig. 1A) (*11*). Surface roughness analysis using atomic force microscopy (AFM) showed that the shiny patches have a striated nanoscale surface topography similar to that of the more continuous surfaces formed at coseismic slip rates (Fig. 1B) (*12*). This suggests

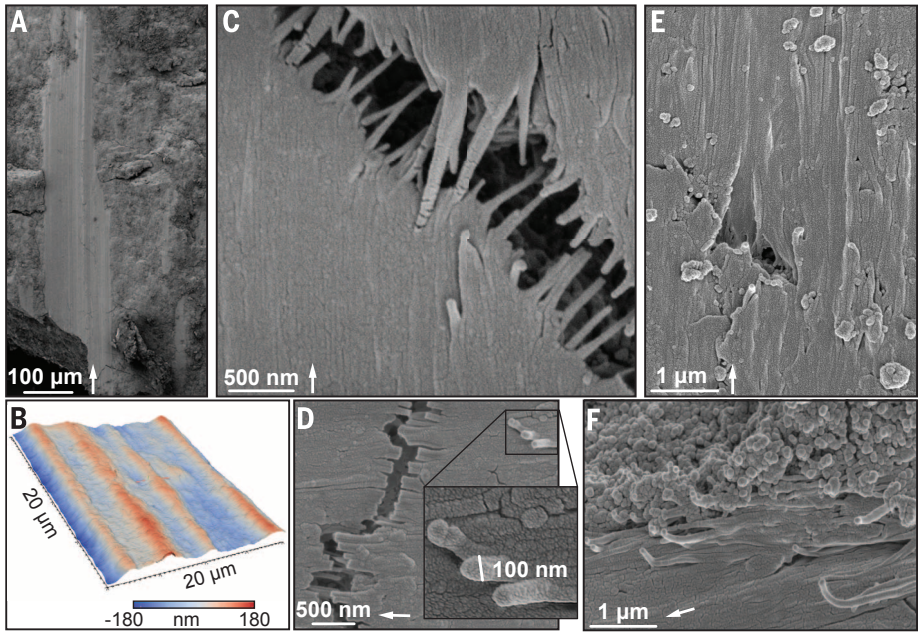


Fig. 1. Mirror-slip surfaces formed at subseismic sliding velocities (1 μm/s). SEM micrographs are secondary electron images of the shear plane. Arrows indicate the shear sense of material above the shear plane. (**A**) Shiny patches formed under v-strengthening conditions (18°C). (**B**) AFM deflection image showing nanoscale surface roughness. (**C**) and (**D**) Nanofibers formed at 18°C. The inset in (D) shows a protruding fiber with a spherule at the tip. (**E**) Part of a shiny patch formed under v-weakening conditions (140°C). (**F**) Chaining of nanospherules at the edge of a fibrous surface. Note the gradation from the zone dominated by nanospherules into the fibrous zone.

[1]Department of Earth Sciences, Utrecht University, Budapestlaan 4, 3584 CD, Utrecht, Netherlands.
[2]Department of Biology, Utrecht University, Padualaan 8, 3584 CH, Utrecht, Netherlands.
*Corresponding author. E-mail: b.a.verberne@uu.nl

Downloaded from www.sciencemag.org on December 14, 2014

that co- and subseismic mirror-slip surfaces share a common origin, independently of their extent or of shearing velocity.

We further investigated the surface morphology of the shiny patches formed in our experiments, using high-resolution scanning electron microscopy (SEM). We found them to consist of ultra–fine-grained films composed of ~100–nm-wide fibers aligned subparallel to the shear direction (Fig. 1, C to F, and fig. S3), embedded within granular zones of frequently clumped, ~100-nm spherules, constituting the bulk PSZ (fig. S4). Individual spherules are sometimes visible within the fibers, especially at their tips (Fig. 1D and fig. S4), whereas linear arrays of nanospherules grading into fibers are also evident (Fig. 1F and fig. S4). On this basis, we suggest that the fibers form by progressive nanospherule chaining. In shiny patches formed in *v*-strengthening samples (18°C), microcracks cutting the nanocoatings show marked extension and plastic bending of the fibers (Fig. 1, C and D). In view of the high confining stress used in our tests (50 MPa), such crack opening could only have occurred upon sample depressurization or disassembly at room conditions (fig. S3). The nanofibers affected often show extreme local extension without the formation of a neck before fracture (Fig. 1, C and D), which are classical characteristics of superplastic deformation (*16*). Shiny patches from *v*-weakening samples (140°C) also consist of ultra–fine-grained fibrous coatings (Fig. 1E) and show essentially the same SEM microstructures as the *v*-strengthening samples (18°C).

We used a focused ion beam (FIB) scanning electron microscope to mill cross sections normal to the shear plane and parallel to the shear direction, to investigate the internal structure of the PSZ and shiny patches (figs. S5 and S6). This showed that the PSZ is a 10- to 100-μm-wide, relatively porous (20 to 30%), nanogranular volume (*10*, *11*), whereas the shiny patches consist of dense, 0.1- to 1-μm-thick, planar fibrous nanocoatings. Some samples show small amounts of an amorphous carbon-rich phase, locally occupying pores ~0.1 to 5 μm in diameter (fig. S5). These deposits show no correlation with mechanical behavior or microstructure, and no temperature changes were measured during shear. We therefore infer them to be derived from the polymer sleeve used to jacket our sample assembly rather than from mechanochemical decomposition (*19*). To resolve the internal structure of the fibrous coatings, we isolated a single nanofiber and milled a thin foil parallel to the fiber length for investigation using transmission electron microscopy (TEM) (Fig. 2A and fig. S7). Selected-area electron diffraction, applied to the nanofiber using a 300-nm aperture, showed ring-shaped diffraction patterns with arcs of greater intensity (Fig. 2B) that attest to a preferred orientation of the (104) or r-plane of calcite, consistent with the optical CPO previously reported for the nanocrystalline PSZs formed in experiments on simulated calcite gouge under the same conditions (supplementary text) (*10*). Bright-field and high-resolution TEM further revealed the presence of spherule-like domains, 100 to 200 nm in diameter, within the nanofiber. These are composed of 5- to 20-nm crystallites, with mutual misorientations of typically just 5° to 20° (Fig. 2, A and C).

All samples investigated using SEM showed evidence for neck growth (sintering) between nanospherules and within nanofibers (Fig. 1D and figs. S4 to S6), implying that diffusive mass transport occurred throughout the PSZ. At the same time, from the porous spherular grain structure observed in the body of the nanogranular PSZ (Fig. 1F and figs. S4 to S6), dilatant granular flow involving spherule rolling and neighbor swapping (Fig. 3A) must also have occurred, whereas the fibrous films suggest cooperative sliding of sintered linear chains of nanospherules (Figs. 1, C to F, and 3, B and C). The Niemeijer-Spiers mechanism of granular flow (*17*) involves competition between dilatation caused by grain-neighbor swapping and compaction by diffusive mass transfer. The original model considers diffusive mass transfer to occur through thin grain boundary fluid films, but it is also valid for solid-state diffusion. For this mechanism to produce *v*-weakening, or a transition thereto, the diffusion-controlled compaction strain rate ($\dot{\varepsilon}_{cp}$) must be of similar order to the granular dilatation rate ($\dot{\varepsilon}_{gr}$). To test whether this is the case for our samples, we estimated the dilatation rate from the shear rate $\dot{\gamma}$ imposed on the PSZ by assuming a dilatation angle of ~35° for granular flow [i.e., using $\dot{\varepsilon}_{gr} \approx 0.7\dot{\gamma}$ (supplementary text) (*17*)] and we compared $\dot{\varepsilon}_{gr}$ with the compaction rate $\dot{\varepsilon}_{cp}$ predicted by equations describing compaction of wet calcite by diffusive mass transport (*20*, *21*). We chose wet calcite because thermogravimetric analyses on nominally dry nanocrystalline calcite powder (particle size ~80 nm) demonstrates a water content of 2 to 3 weight % (fig. S8). This implies that an adsorbed water film a few nanometers thick must coat each ~100-nm particle in the PSZ, which is similar to the amount needed for water-assisted diffusive mass transfer to operate (*22*). Taking the width of the slipping zone to be 1 to 50 μm, $\dot{\varepsilon}_{gr}$ in the shear bands formed in our experiments is on the order of $[10^{-2} - 10^{0}]s^{-1}$, whereas, for a particle size of 100 nm, compaction creep is predicted to occur at a rate $\dot{\varepsilon}_{cp} \approx [10^{-2} - 10^{0}]s^{-1} \approx \dot{\varepsilon}_{gr}$ (*18*). This confirms that the Niemeijer-Spiers model is capable of explaining the *v*-weakening behavior seen in our experiments at temperatures above 80° to 100°C (fig. S2) (*9*–*11*), suggesting that this may be caused by enhanced compaction rates due to faster diffusive transport at these temperatures. Application of the model assuming solid-state grain boundary diffusion (*23*) demonstrates that this is 10 to 20 orders of

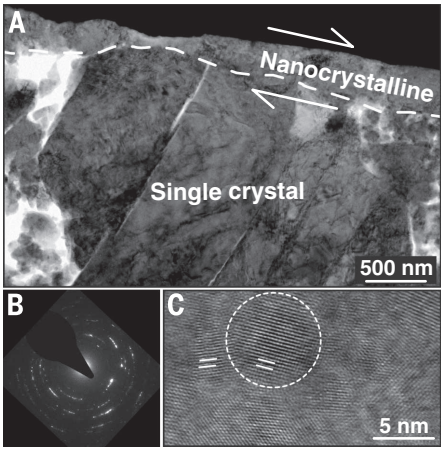


Fig. 2. Individual nanofiber imaged using TEM. (**A**) Bright-field electron micrograph of a single nanofiber recovered from a mirror-slip surface using FIB-SEM (fig. S7). (**B**) Selected-area electron diffraction pattern showing arcs of greater intensity in the diffraction ring corresponding with the (104) or r-plane of calcite (supplementary text). (**C**) High-resolution micrograph taken from a nanofiber. The dashed white circle highlights a single nanocrystallite, and the thick white lines indicate the orientation of the (104) plane (supplementary text).

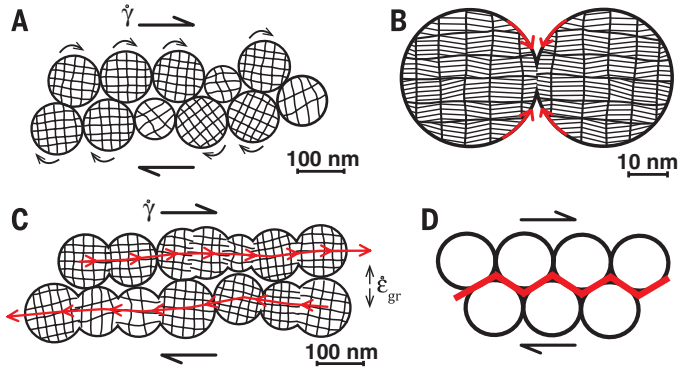


Fig. 3. Two-dimensional representation of envisaged mechanism of nanofiber formation during nanogranular flow with diffusive mass transfer. (**A**) Nanospherule rolling and grain-neighbor swapping with (**B**) local sintering attachment (neck growth) at preferred high-energy interfaces, leading to (**C**) formation of cooperatively sliding nanofibers (linear arrays or chains of sintered nanospherules). Note the cavitation (porosity development) in (A) to (C), compared with (**D**), which represents a zero-porosity grain boundary sliding mechanism that is fully diffusion-accommodated (*30*).

magnitude too slow to produce v-weakening, except perhaps at temperatures of 500° to 700°C reached in experiments at coseismic slip velocities (*4*, *5*, *14*, *19*).

The internal polycrystalline substructure of the nanospherules and nanofibers that we observed bears a striking similarity to microstructures found in shocked ductile metals (*24*). As in metals, the well-known ductility of calcite (*25*) may allow the ~5- to 20-nm substructure to form by progressive development of nano–cell walls from dense dislocation networks generated by crystal plasticity. Plastic deformation, fracturing, and abrasion presumably generated the observed nanospherules from the starting "gouge." To explain the chaining of nanospherules, producing the observed fiber structure and CPO, we note that oriented attachment at coherent nanoparticle interfaces is widely reported as a mechanism by which nanocrystallites can rapidly coalesce to form single crystals (*26*, *27*), also in calcite (*28*). On this basis, we suggest that the strong anisotropy in the surface energy of calcite produced similar preferred sintering (neck growth) at high-energy crystallographic interfaces between neighboring spherules (Fig. 3, B and C), leading to dynamic chaining and alignment of the lowest-energy (104) plane (*29*) parallel to the shear plane, and thus to the observed fibrous structure and CPO.

In principle, the mechanism of frictional slip that we propose (Fig. 3, A to C, and fig. S1) is similar to the Ashby-Verrall model for superplasticity by diffusion-accommodated grain boundary sliding (GBS) (Fig. 3D) (*30*), but allows for frictional GBS and for intergranular cavitation (porosity generation by dilatation) when diffusive mass transport is too slow to accommodate GBS. Our findings imply that nanocrystalline PSZs developed in calcite faults can produce v-weakening, and hence seismogenic fault friction, by a mechanism of cooperative nanogranular or nanofiber flow plus diffusive mass transfer (Fig. 3, A to C), even in the upper crust where temperatures are generally considered too low to support diffusion or superplasticity at active fault slip rates. The reason that these processes are observed in our experiments is because diffusive mass transfer is dramatically accelerated by the nanogranular nature of the slip-zone rock that forms, and by water-enhanced grain boundary diffusion. A similar mechanism can also be envisaged to operate at coseismic slip rates, where the high temperatures generated will promote solid-state diffusion. Given the abundant recent observations of nanogranular fault surfaces in tectonically active terrains (*1*–*7*), and the anomalously high rates of diffusion found in nanomaterials (*15*, *16*), the proposed mechanism may be relevant not only to faults cutting calcite-rich rocks such as limestones, but to crustal seismogenesis in general.

REFERENCES AND NOTES

1. J. S. Chester, F. M. Chester, A. K. Kronenberg, *Nature* **437**, 133–136 (2005).
2. K.-F. Ma *et al.*, *Nature* **444**, 473–476 (2006).
3. S. Siman-Tov, E. Aharonov, A. Sagy, S. Emmanuel, *Geology* **41**, 703–706 (2013).
4. S. A. F. Smith *et al.*, *Geology* **41**, 63–66 (2013).
5. M. Fondriest *et al.*, *Geology* **41**, 1175–1178 (2013).
6. J. D. Kirkpatrick, C. D. Rowe, J. C. White, E. E. Brodsky, *Geology* **41**, 1015–1018 (2013).
7. J. P. Evans, M. R. Prante, S. U. Janecke, A. K. Ault, D. L. Newell, *Geology* **42**, 623–626 (2014).
8. C. H. Scholz, *Nature* **391**, 37–42 (1998).
9. B. A. Verberne, C. He, C. J. Spiers, *Bull. Seismol. Soc. Am.* **100**, 2767–2790 (2010).
10. B. A. Verberne *et al.*, *Geology* **41**, 863–866 (2013).
11. B. A. Verberne *et al.*, *Pure Appl. Geophys.* **171**, 2617–2640 (2014).
12. X. Chen, A. S. Madden, B. R. Bickmore, Z. Reches, *Geology* **41**, 739–742 (2013).
13. R. Han, T. Hirose, T. Shimamoto, Y. Lee, J.-i. Ando, *Geology* **39**, 599–602 (2011).
14. G. Di Toro *et al.*, *Nature* **471**, 494–498 (2011).
15. R. Würschum, S. Herth, U. Brossmann, *Adv. Eng. Mater.* **5**, 365–372 (2003).
16. S. C. Tjong, H. Chen, *Mater. Sci. Eng. Rep.* **45**, 1–88 (2004).
17. A. R. Niemeijer, C. J. Spiers, *J. Geophys. Res.* **112**, B10405 (2007).
18. Materials and methods are available as supplementary materials on *Science* Online.
19. K. Oohashi *et al.*, *Geology* **42**, 787–790 (2014).
20. C. J. Spiers, S. De Meer, A. R. Niemeijer, X. Zhang, in *Physicochemistry of Water in Geological and Biological Systems – Structures and Properties of Thin Aqueous Films*, S. Nakashima, C. J. Spiers, L. Mercury, P. A. Fenter, M. F. Hochella Jr, Eds. (Universal Academy Press, Tokyo, 2004), pp. 129–158.
21. X. Zhang, C. J. Spiers, C. J. Peach, *J. Geophys. Res.* **115**, B09217 (2010).
22. J.-P. Gratier, D. K. Dysthe, F. Renard, *Adv. Geophys.* **54**, 47–179 (2013).
23. J. R. Farver, R. A. Yund, *Contrib. Mineral. Petrol.* **123**, 77–91 (1996).
24. N. R. Tao *et al.*, *Acta Mater.* **50**, 4603–4616 (2002).
25. L. A. Kennedy, J. C. White, *Geology* **29**, 1027–1030 (2001).
26. R. L. Penn, J. F. Banfield, *Science* **281**, 969–971 (1998).
27. H. Zhang, J. F. Banfield, *CrystEngComm* **16**, 1568–1578 (2014).
28. N. Gehrke, H. Cölfen, N. Pinna, M. Antonietti, N. Nassif, *Cryst. Growth Des.* **5**, 1317–1319 (2005).
29. N. H. de Leeuw, S. C. Parker, *J. Chem. Soc. Faraday Trans.* **93**, 467–475 (1997).
30. M. F. Ashby, R. A. Verrall, *Acta Metall. Mater.* **21**, 149–163 (1973).

ACKNOWLEDGMENTS

We thank A. Niemeijer, J. Chen, V. Toy, and H. de Bresser. H. King is thanked for the AFM measurements and P. van Krieken for the thermogravimetric analysis. B.A.V. was supported by grant 2011-75, awarded by the Netherlands Research Centre for Integrated Solid Earth Sciences; O.P. by Veni grant 863.13.006, awarded by the Netherlands Organisation for Scientific Research (NWO); and D.A.M.D.W. by ISES grant 2011-74. NWO funded the FIB-SEM. All data are available in the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/346/6215/1342/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S8
Tables S1 to S2
References (*31*–*40*)

21 July 2014; accepted 11 November 2014
10.1126/science.1259003

SOLID STATE THEORY

Quantum spin Hall effect in two-dimensional transition metal dichalcogenides

Xiaofeng Qian,[1]* Junwei Liu,[2]* Liang Fu,[2]† Ju Li[1]†

Quantum spin Hall (QSH) effect materials feature edge states that are topologically protected from backscattering. However, the small band gap in materials that have been identified as QSH insulators limits applications. We use first-principles calculations to predict a class of large-gap QSH insulators in two-dimensional transition metal dichalcogenides with 1T′ structure, namely, 1T′-MX_2 with M = (tungsten or molybdenum) and X = (tellurium, selenium, or sulfur). A structural distortion causes an intrinsic band inversion between chalcogenide-*p* and metal-*d* bands. Additionally, spin-orbit coupling opens a gap that is tunable by vertical electric field and strain. We propose a topological field effect transistor made of van der Waals heterostructures of 1T′-MX_2 and two-dimensional dielectric layers that can be rapidly switched off by electric field through a topological phase transition instead of carrier depletion.

The discovery of graphene (*1*) has fuelled vigorous investigation of two-dimensional (2D) materials (*2*), revealing a wide range of extraordinary properties (*3*–*5*) and functionalities (*6*, *7*). Owing to their atomic thickness, 2D materials can be horizontally patterned through chemical and mechanical techniques (*8*). Moreover, the weak van der Waals (vdW) interaction between adjacent layers enables vertical stacking of different 2D materials, forming vdW heterostructures (*9*), which offer unprecedented opportunities for exploring quantum electronics at the nanoscale.

[1]Department of Nuclear Science and Engineering and Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
[2]Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.
*These authors contributed equally to this work. **†Corresponding author. E-mail: liangfu@mit.edu (L.F.); liju@mit.edu (J.L.)**

Downloaded from www.sciencemag.org on December 14, 2014

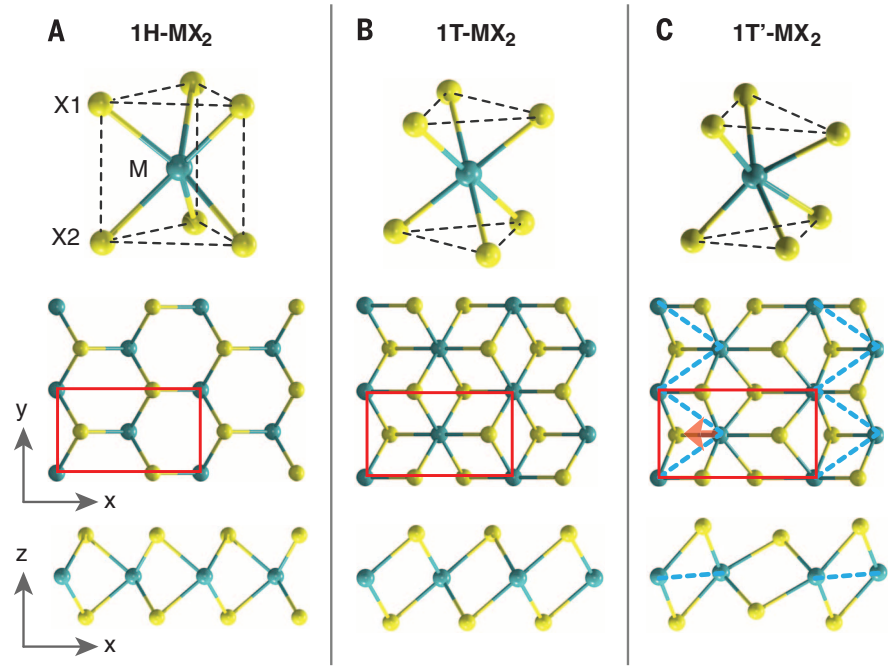


Fig. 1. Atomistic structures of monolayer transition metal dichalcogenides MX_2. M stands for (W, Mo) and X stands for (Te, Se, S). (**A**) 1H-MX_2 in ABA stacking with $P\bar{6}m2$ space group. (**B**) 1T-MX_2 in ABC stacking with $P\bar{3}m2$ space group. (**C**) 1T′-MX_2, distorted 1T-MX_2, where the distorted M atoms form 1D zigzag chains indicated by the dashed blue line. The unit cell is indicated by red rectangles.

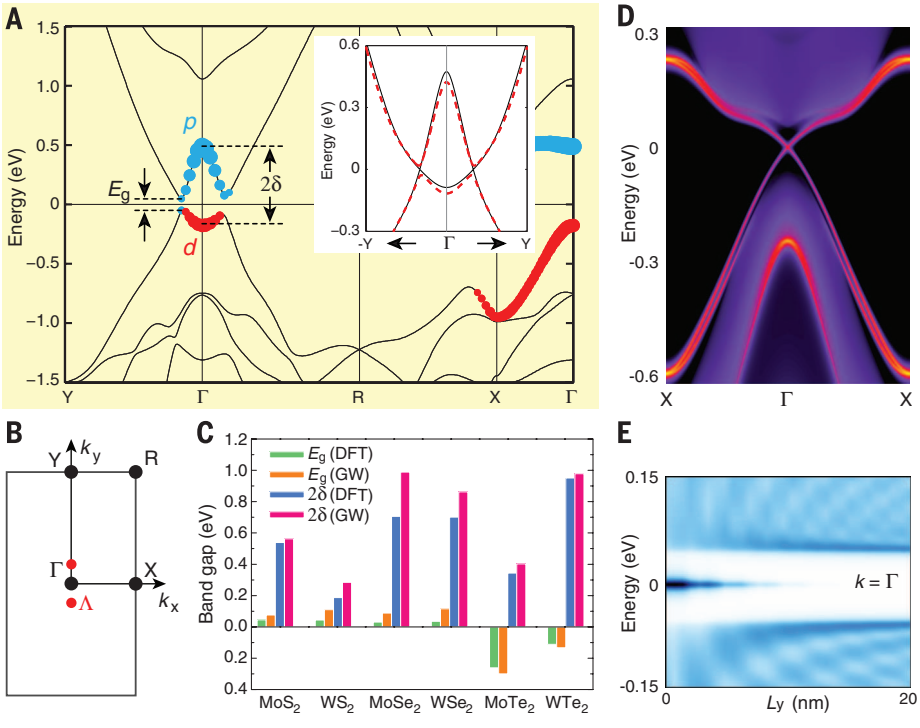


Fig. 2. Calculated electronic structures of 1T′-MX_2. (**A**) Band structure and (**B**) Brillouin zone of 1T′-MoS_2. E_g, fundamental gap; 2δ, inverted gap. The inset compares band structures with (red dashed line) and without (black solid line) spin-orbit coupling. Four time-reversal invariant momenta are marked by black dots and labeled as Γ, X, Y, and R. The locations of the fundamental gap are marked by red dots and labeled by Λ. (**C**) Fundamental gap (E_g) and inverted gap (2δ) of all six 1T′-MX_2. Detailed values are listed in table S1. (**D**) Edge density of states and (**E**) local density of states at Γ point as a function of distance away from the edge (L_y). A sharp peak from edge states appears in the gap with a decay length of ~5 nm.

Quantum spin Hall (QSH) insulators (*10*–*16*) have an insulating bulk but conducting edge states that are topologically protected from backscattering by time-reversal symmetry. Quantized conductance through QSH edge states have been experimentally demonstrated in HgTe/CdTe (*13*, *14*) and InAs/GaSb (*17*, *18*) quantum wells. This could in principle provide an alternative route to quantum electronic devices with low dissipation. However, the realization of such QSH-based devices for practical applications is impeded by three critical factors: (i) band gaps of existing QSH insulators are too small, which limits the operating regime to low temperatures. This has motivated efforts to search for large-gap QSH insulators (*19*–*26*); (ii) the small number of conducting channels (e^2/h per edge, where e is the elementary charge and h is Planck's constant) results in a small signal-to-noise ratio; and (iii) efficient methods of fast on/off switching are lacking.

Here, we use first-principles calculations to show that 2D materials can provide a practical platform for developing topological electronic devices that may potentially overcome the above hurdles. Specifically, we predict a class of large-gap (~0.1 eV) QSH insulators in 2D transition metal dichalcogenides (TMDCs) MX_2 with M = (W, Mo) and X = (Te, Se, S). We demonstrate the feasibility of a vdW-heterostructured topological field-effect transistor (vdW-TFET) made of 2D atomic layer materials. The proposed device exhibits a parametrically enhanced conductance through QSH edge channels in the "on" state and can be rapidly switched off via a topological phase transition by applying a vertical electric field. Our findings have potential applications in low-power quantum electronics and spintronics and may enable topological quantum computing based on Majorana fermions (*27*–*30*).

Monolayer TMDCs (*31*)—MX_2 with M = (W, Mo) and X = (Te, Se, S)—possess a variety of polytypic structures such as 1H, 1T, and 1T′ (*32*, *33*) shown in Fig. 1. The most-studied 1H structure is a sandwich of three planes of 2D hexagonally packed atoms, X-M-X, forming Bernal (ABA) stacking. In contrast, the three atomic planes in the 1T structure form rhombohedral (ABC) stacking. It has been known that the 1T structure in MX_2 is typically unstable in free-standing condition and undergoes a spontaneous lattice distortion in the x direction to form a period-doubling 2 × 1 distorted structure—i.e., the 1T′ structure, consisting of 1D zigzag chains along the y direction (Fig. 1C) (*33*); this structure is the subject of our work.

The electronic structures of various 1T′-MX_2 materials were obtained by many-body perturbation theory calculations (*34*). Figure 2A shows a typical band structure of 1T′-MX_2 using 1T′-MoS_2 as an example, and the results of the other five compounds are shown in fig. S1. Unlike its 1H or 1T counterparts, 1T′-MoS_2 is a semiconductor with a fundamental gap (E_g) of about 0.08 eV, located at $\Lambda = \pm(0, 0.146)\text{Å}^{-1}$

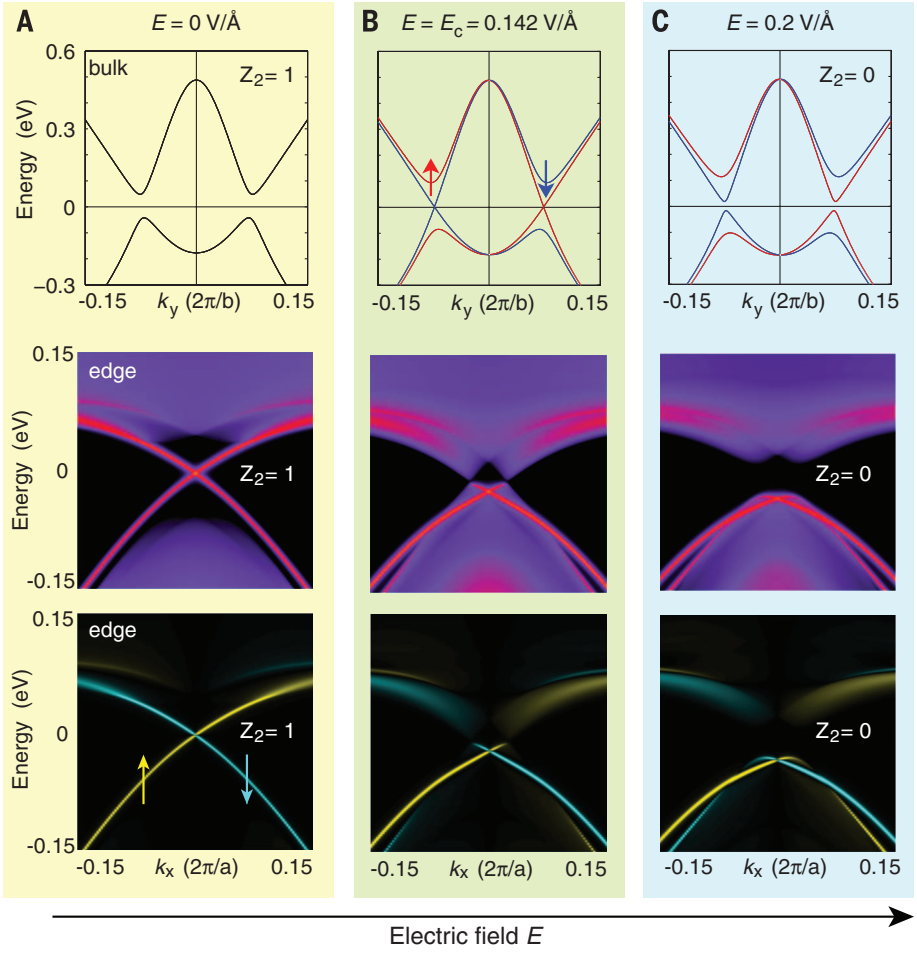


Fig. 3. Vertical electric field induced a topological phase transition. Shown are calculations for monolayer 1T′-MoS_2 under electric fields of (**A**) 0, (**B**) 0.142, and (**C**) 0.2 V/Å, respectively. The first, second, and third rows show the bulk band structure, edge density of states, and edge spin polarization, respectively.

(red dots in Fig. 2B). The conduction and valence bands display a camelback shape near Γ in the 2D Brillouin zone (BZ) (see Fig. 2B), suggestive of band inversion with a large inverted gap (2δ) at Γ of about 0.6 eV. The fundamental gap (E_g) and inverted gap (2δ) of all six 1T′-MX_2 are shown in Fig. 2C and listed in table S1. The inverted gap (2δ) was later used in the fitting of a low-energy $\boldsymbol{k} \cdot \boldsymbol{p}$ model (*34*). Because the 1T′ structure has inversion symmetry, we calculated the Z_2 index (0 or 1, indicating trivial or nontrivial topology, respectively) based on the parity criterion (*35*) of valence bands (fig. S2) and found Z_2 = 1, indicating that 2D 1T′-MX_2 materials are in the QSH insulator phase.

To understand the origin of the inverted band structure near Γ, we analyzed the orbital character of the bands and found that the valence band mainly consists of metal *d*-orbitals and the conduction band mainly consists of chalcogenide *p* orbitals, as expected. The unexpected band inversion arises from the period doubling of the metal chain in the 1T′ structure, which lowers the metal *d* orbital below chalcogenide *p* orbital (Fig. 2A) around Γ. The large band inversion at Γ (~0.6 eV) is an intrinsic characteristic of the 1T′ structure and already takes place without including spin-orbit interaction in the first-principles calculations. In the absence of spin-orbit coupling, this band inversion leads to the appearance of two Dirac cones centered at finite momenta on Y-Γ-Y in 2D BZ. Spin-orbit coupling then opens up a fundamental gap of 0.08 eV at the Dirac points (Fig. 2A, inset, and fig. S3), leading to a QSH insulator through a similar mechanism as the Kane-Mele model for graphene (*10*); we show that this is the case by using a low-energy $\boldsymbol{k} \cdot \boldsymbol{p}$ Hamiltonian (*34*) (fig. S4). In addition to MoS_2, we found that all other five 1T′-MX_2 have Z_2 nontrivial band topology.

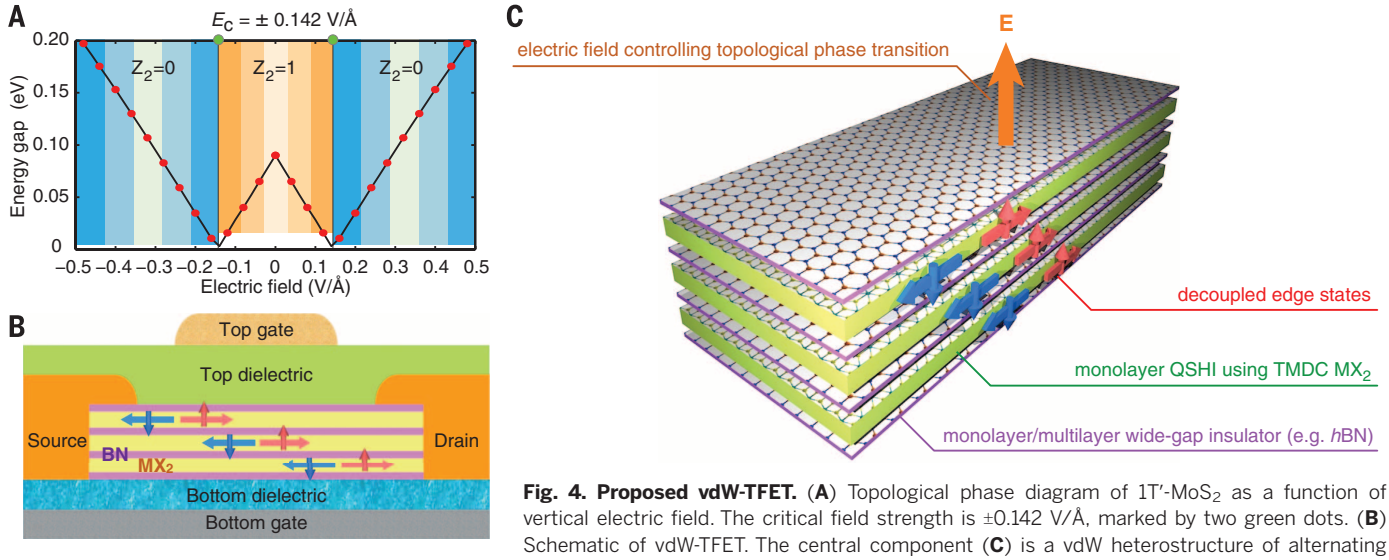


Fig. 4. Proposed vdW-TFET. (**A**) Topological phase diagram of 1T′-MoS_2 as a function of vertical electric field. The critical field strength is ±0.142 V/Å, marked by two green dots. (**B**) Schematic of vdW-TFET. The central component (**C**) is a vdW heterostructure of alternating monolayer 1T′-MX_2 and mono-/multilayer wide-gap insulators such as hBN, with horizontal width as narrow as ~20 nm. Carriers (charge or spin) are injected from the source electrode and ejected into the drain electrode. On/off switch is controlled by vertical electric field through the top and bottom gates. Mono-/multilayer wide-gap insulators effectively screen the interaction between adjacent MX_2 layers, preventing them from detrimental topological phase change and parametrically increasing the number of edge channels.

The QSH insulator phase in 2D 1T′-MX_2 leads to helical edge states that are protected from localization and elastic backscattering by time-reversal symmetry. Figure 2D displays the edge states of 1T′-MoS_2 obtained from surface Green's function calculations (*34*), showing a pair of counter-propagating edge modes with a high velocity of ~1.0×10^5 m/s that cross at Γ and carry opposite spin polarizations. From the local density of states in Fig. 2E, we find the decay length of these helical edge states to be as short as 5 nm. The short localization length can greatly reduce scattering with bulk states (*36*), hence increasing the transport lifetime. Similar results were found for five other TMDCs (figs. S5 and S6).

The inverted bands between chalcogenide's *p* and metal's *d* orbitals located on well-separated planes offer a facile mechanism to control topological electronic properties by electric field, which is highly desirable for vdW devices. Based on first-principles calculations, we found that a vertical electric field indeed induces a topological phase transition in 1T′-MX_2, as shown in Fig. 3 for 1T′-MoS_2. Here, the electric field breaks inversion symmetry and introduces a strong Rashba splitting of the doubly degenerate bands near the fundamental gap E_g at the Λ points. As the field increases, E_g first decreases to zero at a critical field strength of 0.142 V/Å and then reopens (see Fig. 4A). This gap-closing transition induces a topology change to a trivial phase (*11*, *37*), leading to the destruction of helical edge states.

The above findings point to the possibility of an all-electrical control of the on/off charge/spin conductance of helical edge states, which would have substantial implications for QSH-based devices (*38*–*43*). Here, we propose a TFET based on vdW heterostructures of 2D TMDCs and 2D wide-gap insulators. The proposed device is sketched in Fig. 4B, where the top and bottom gates supply the vertical electric field to control the on/off function. Two-dimensional wide-gap insulators such as hexagonal boron nitride (hBN) (*44*) electrically insulate adjacent QSH layers, hence protecting parallel helical edge channels from being gapped by interlayer hybridization. This can parametrically increase the number of edge transport channels. Under ideal conditions, this device will support dissipationless charge/spin transport in the "on" state (Z_2 = 1), with a quantized conductance of $2Ne^2/h$, where N is the number of QSH layers. Applying a moderate electric field will transform 1T′-MX_2 into an ordinary insulator (Z_2 = 0) and turn the edge conduction off (Fig. 4A).

The proposed vdW-TFET may possess important technical advantages, thanks to an operation mechanism that is fundamentally different from traditional metal-oxide-semiconductor FETs (MOS-FET). MOS-FET works by injecting/depleting free carrier beneath the gate, with a resistor–capacitor-type response time influenced by carrier concentration and mobility in the working channel, whereas our vdW-TFET works by switching the presence/absence of topologically protected conduction channels. The electric field–induced topological phase transition can happen very rapidly, with an electronic response time scale (*45*, *46*). In addition, the vdW heterostructure and the 2D nature of vdW-TFET make it convenient to both vertically and horizontally pattern devices at a large scale. The short decay length of helical edge states allows the minimum horizontal feature size of vdW-TFET to be ~20 nm, enabling high device density on a chip. Horizontally patterned TFET devices of such nanoscale size have large edge-to-area ratios, which will greatly reduce the contribution of thermally populated bulk carriers to the total electric current and hence enhance the on/off ratio. The predicted QSH insulators and their vdW heterostructures may provide a platform for realizing low-dissipation quantum electronics and spintronics (*38*–*43*).

From a materials perspective, the 1T′ structure in monolayer MoS_2 has recently been observed in high-resolution transmission electron microscopy experiments (*33*). We have checked the structural stability of 1T′ by carrying out first-principles calculations (fig. S7). Although MoS_2, $MoSe_2$, $MoTe_2$, WS_2, and WSe_2 are more stable in the 1H structure, we find that there exists a large energy barrier of more than 1 eV/MX_2 between 1T′ and 1H for all MX_2, suggesting that the 1T′ structure can be stabilized under appropriate chemical (*47*), thermal, or mechanical conditions. Moreover, we found that the 1T structure is unstable and will spontaneously relax to 1T′, like in a Peierls distortion. We also verified the stability of the 1T′ structure by computing its phonon dispersion (fig. S8). The absence of imaginary frequencies throughout the 2D BZ provides direct evidence of the structural stability in all 1T′-MX_2.

Encouragingly, the desired 1T′ structure with intrinsic band inversion is the natural ground-state structure of WTe_2 (*31*), also confirmed by our stability calculation. Although monolayer WTe_2 was found to be a semimetal from our calculation, the small negative fundamental gap can be easily lifted by a tensile strain as small as ~1% (fig. S9). We also found that hBN used as a dielectric in the aforementioned vdW-TFET device has little effect on the electronic structure of 1T′-WTe_2 near the Fermi level (fig. S10). These attractive features make WTe_2 potentially the most promising 2D TMDC for realizing the QSH effect and related vdW devices.

REFERENCES AND NOTES

1. K. S. Novoselov *et al*., *Science* **306**, 666–669 (2004).
2. K. S. Novoselov *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **102**, 10451–10453 (2005).
3. C. Lee, X. Wei, J. W. Kysar, J. Hone, *Science* **321**, 385–388 (2008).
4. A. H. Castro Neto, F. Guinea, N. M. R. Peres, K. S. Novoselov, A. K. Geim, *Rev. Mod. Phys.* **81**, 109–162 (2009).
5. K. F. Mak, C. Lee, J. Hone, J. Shan, T. F. Heinz, *Phys. Rev. Lett.* **105**, 136805 (2010).
6. Q. H. Wang, K. Kalantar-Zadeh, A. Kis, J. N. Coleman, M. S. Strano, *Nat. Nanotechnol.* **7**, 699–712 (2012).
7. J. Feng, X. Qian, C.-W. Huang, J. Li, *Nat. Photonics* **6**, 866–872 (2012).
8. K. S. Kim *et al*., *Nature* **457**, 706–710 (2009).
9. A. K. Geim, I. V. Grigorieva, *Nature* **499**, 419–425 (2013).
10. C. L. Kane, E. J. Mele, *Phys. Rev. Lett.* **95**, 226801 (2005).
11. C. L. Kane, E. J. Mele, *Phys. Rev. Lett.* **95**, 146802 (2005).
12. B. A. Bernevig, S.-C. Zhang, *Phys. Rev. Lett.* **96**, 106802 (2006).
13. B. A. Bernevig, T. L. Hughes, S.-C. Zhang, *Science* **314**, 1757–1761 (2006).
14. M. König *et al*., *Science* **318**, 766–770 (2007).
15. M. Z. Hasan, C. L. Kane, *Rev. Mod. Phys.* **82**, 3045–3067 (2010).
16. X.-L. Qi, S.-C. Zhang, *Rev. Mod. Phys.* **83**, 1057–1110 (2011).
17. C. Liu, T. L. Hughes, X.-L. Qi, K. Wang, S.-C. Zhang, *Phys. Rev. Lett.* **100**, 236601 (2008).
18. L. Du, I. Knez, G. Sullivan, R.-R. Du, http://arxiv.org/abs/1306.1925 (2013).
19. S. Murakami, *Phys. Rev. Lett.* **97**, 236805 (2006).
20. C.-C. Liu, W. Feng, Y. Yao, *Phys. Rev. Lett.* **107**, 076802 (2011).
21. D. Xiao, W. Zhu, Y. Ran, N. Nagaosa, S. Okamoto, *Nat. Commun.* **2**, 596–596 (2011).
22. Y. Xu *et al*., *Phys. Rev. Lett.* **111**, 136804 (2013).
23. H. Weng, X. Dai, Z. Fang, *Phys. Rev. X* **4**, 011002 (2014).
24. M. A. Cazalilla, H. Ochoa, F. Guinea, *Phys. Rev. Lett.* **113**, 077201 (2014).
25. D. Kong, Y. Cui, *Nat. Chem.* **3**, 845–849 (2011).
26. K. Yang, W. Setyawan, S. Wang, M. Buongiorno Nardelli, S. Curtarolo, *Nat. Mater.* **11**, 614–619 (2012).
27. L. Fu, C. L. Kane, *Phys. Rev. Lett.* **100**, 096407 (2008).
28. L. Fu, C. L. Kane, *Phys. Rev. B* **79**, 161408 (2009).
29. J. Alicea, *Rep. Prog. Phys.* **75**, 076501 (2012).
30. S. Mi, D. I. Pikulin, M. Wimmer, C. W. J. Beenakker, *Phys. Rev. B* **87**, 241405 (2013).
31. J. A. Wilson, A. D. Yoffe, *Adv. Phys.* **18**, 193–335 (1969).
32. J. Heising, M. G. Kanatzidis, *J. Am. Chem. Soc.* **121**, 11720–11732 (1999).
33. G. Eda *et al*., *ACS Nano* **6**, 7311–7317 (2012).
34. Supplementary materials are available on *Science* Online.
35. L. Fu, C. L. Kane, *Phys. Rev. B* **76**, 045302 (2007).
36. J. I. Väyrynen, M. Goldstein, L. I. Glazman, *Phys. Rev. Lett.* **110**, 216402 (2013).
37. S. Murakami, *New J. Phys.* **9**, 356 (2007).
38. J. Wunderlich *et al*., *Science* **330**, 1801–1804 (2010).
39. D. Pesin, A. H. MacDonald, *Nat. Mater.* **11**, 409–416 (2012).
40. F. Mahfouzi, N. Nagaosa, B. K. Nikolić, *Phys. Rev. Lett.* **109**, 166602 (2012).
41. C. Brüne *et al*., *Nat. Phys.* **8**, 485–490 (2012).
42. C. H. Li *et al*., *Nat. Nanotechnol.* **9**, 218–224 (2014).
43. A. R. Mellnik *et al*., *Nature* **511**, 449–451 (2014).
44. C. R. Dean *et al*., *Nat. Nanotechnol.* **5**, 722–726 (2010).
45. J. Liu *et al*., *Nat. Mater.* **13**, 178–183 (2014).
46. P. Michetti, B. Trauzettel, *Appl. Phys. Lett.* **102**, 063503 (2013).
47. M. Calandra, *Phys. Rev. B* **88**, 245428 (2013).

ACKNOWLEDGMENTS

We acknowledge support from NSF under award DMR-1120901 (X.Q. and J.L.), U.S. Department of Energy, Office of Basic Energy Sciences, Division of Materials Sciences and Engineering under award DE-SC0010526 (L.F.), and the Science and Technology Center on Integrated Quantum Materials, NSF grant no. DMR-1231319 (J.W.L.). Computational time on the Extreme Science and Engineering Discovery Environment (XSEDE) under grant nos. TG-DMR130038 and TG-DMR140003 is gratefully acknowledged. X.Q. and J.W.L. contributed equally to this work. MIT has applied for a provisional patent based on the content of this paper. Both atomic and electronic structures of 1T′-MX_2 can be found in the supplementary materials.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/346/6215/1344/suppl/DC1
Supplementary Methods
Figs. S1 to S10
Tables S1 and S2
References (*48*–*62*)

2 June 2014; accepted 5 November 2014
10.1126/science.1256815

ULTRAFAST DYNAMICS

Attosecond band-gap dynamics in silicon

Martin Schultze,[1,2]*† Krupa Ramasesha,[1]* C.D. Pemmaraju,[3] S.A. Sato,[4] D. Whitmore,[1] A. Gandman,[1] James S. Prell,[1] L. J. Borja,[1] D. Prendergast,[3] K. Yabana,[4,5] Daniel M. Neumark,[1,6]† Stephen R. Leone[1,6,7]†

Electron transfer from valence to conduction band states in semiconductors is the basis of modern electronics. Here, attosecond extreme ultraviolet (XUV) spectroscopy is used to resolve this process in silicon in real time. Electrons injected into the conduction band by few-cycle laser pulses alter the silicon XUV absorption spectrum in sharp steps synchronized with the laser electric field oscillations. The observed ~450-attosecond step rise time provides an upper limit for the carrier-induced band-gap reduction and the electron-electron scattering time in the conduction band. This electronic response is separated from the subsequent band-gap modifications due to lattice motion, which occurs on a time scale of 60 ± 10 femtoseconds, characteristic of the fastest optical phonon. Quantum dynamical simulations interpret the carrier injection step as light-field–induced electron tunneling.

Excitation of electrons from valence band (VB) states into mobile conduction band (CB) states by light is one of the most fundamental phenomena in semiconductor physics. The dynamics of these charge carriers have been extensively studied previously on multifemtosecond to microsecond time scales, providing insight into postexcitation processes such as carrier cooling, transport, and electron-hole recombination (*1*). However, the buildup and interaction of carriers in the conduction band during excitation have thus far proven too rapid to probe directly. Interest in the time evolution of the excitation process is motivated by far-reaching implications regarding the ultimate speed of electronics (*2*), solid-state coherent control (*3*, *4*), and the mechanisms of dielectric breakdown (*5*–*7*). In previous experiments, only at very low excitation densities (*8*) or for quantum mechanically suppressed electron-phonon coupling has it been possible to disentangle electronic and lattice dynamics (*9*). Thus, a key question remained open as to whether the reduction of the measured band gap of semiconductors as a function of CB population [for optically (*10*)/thermally (*11*) excited or doped (*12*) samples] involves the rearrangement of the crystal lattice or is mainly a result of carrier-carrier interactions in the CB.

Attosecond solid-state spectroscopy (*13*) in tandem with time-dependent density functional theory (TD-DFT) is used here to follow the initial steps of the excitation mechanism in real time in silicon. Probing the sample with isolated attosecond pulses near the silicon 2*p* core transition ($L_{2,3}$-edge at 99 eV) reveals that electrons excited across the band gap cause an instantaneous reduction of the band-gap energy due to carrier-carrier interactions evolving on an ~450-attosecond time scale. The electronic response is completed before the lattice motion affects the band gap, with a longer time constant of 60 ± 10 fs resembling the period of the fastest optical phonon in silicon of 64 fs (15.6 THz) (*14*, *15*).

The principles of attosecond transient absorption spectroscopy of semiconductors are illustrated in Fig. 1. The band structure of crystalline silicon accommodates both direct and indirect excitations of electrons across the band gap (Fig. 1A). At low intensities, a near-infrared (NIR) pump pulse can in principle promote electrons from the VB maximum across the indirect band gap (E_{gap} = 1.2 eV), satisfying momentum conservation by the addition of a phonon. However, at higher intensities above 10^{10} W/cm^2, the creation of electron hole pairs across the larger direct band gap (E_{gap} = 3.2 eV) becomes possible and dominates the transition rate. This transition occurs without phonon participation, but the larger energy separation of the direct band gap requires the simultaneous absorption of two or more NIR or visible photons (E_{photon} = 1.1 – 2.7 eV). Here, we establish that for light intensities of >10^{12} W/cm^2, electron tunneling driven by the electric field of light is the main excitation pathway.

Attosecond pulses tuned to the silicon $L_{2,3}$ edge centered at 99 eV, as displayed in Fig. 1A, are used to measure the dynamics of electrons injected into the CB by few-cycle NIR laser pump pulses. The probe pulses are shorter than 100 attoseconds, as confirmed by photoelectron streaking spectroscopy [see the supplementary

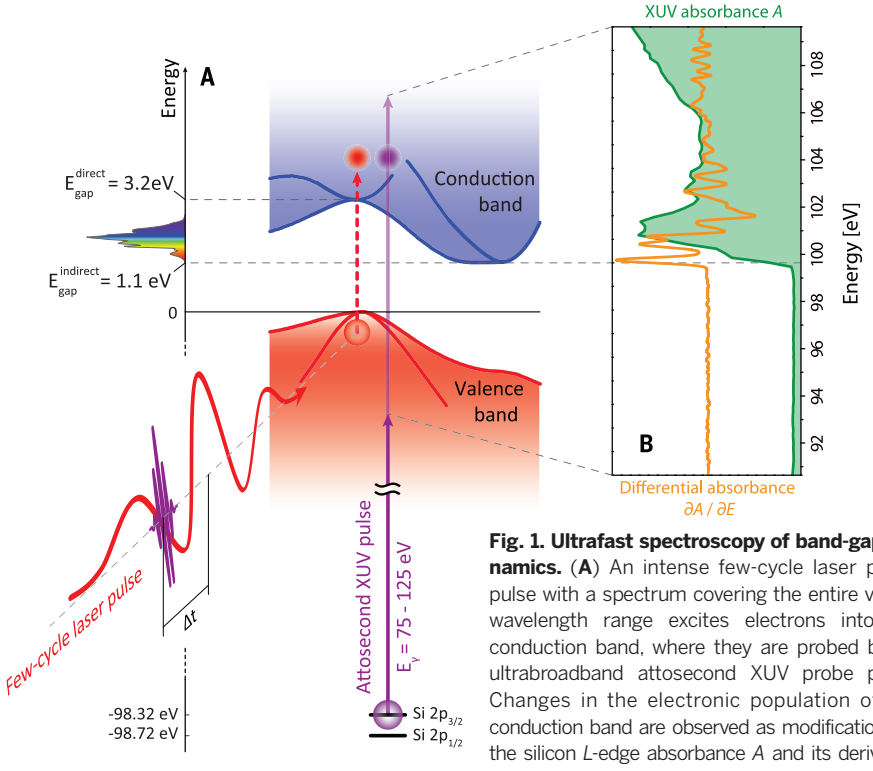


Fig. 1. Ultrafast spectroscopy of band-gap dynamics. (**A**) An intense few-cycle laser pump pulse with a spectrum covering the entire visible wavelength range excites electrons into the conduction band, where they are probed by an ultrabroadband attosecond XUV probe pulse. Changes in the electronic population of the conduction band are observed as modifications of the silicon *L*-edge absorbance *A* and its derivative $\partial A/\partial E$ (**B**). By varying the time delay Δt between the two pulses, dynamic changes to the *L*-edge spectrum can be tracked and translated into the evolution of the conduction band population.

[1]Department of Chemistry, University of California, Berkeley, CA 94720, USA. [2]Fakultät für Physik, Ludwig-Maximilians-Universität, Am Coulombwall 1, D-85748 Garching, Germany. [3]The Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [4]Graduate School of Pure and Applied Sciences, University of Tsukuba, Tsukuba 305-8571, Japan. [5]Center for Computational Sciences, University of Tsukuba, Tsukuba 305-8577, Japan. [6]Chemical Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, USA. [7]Department of Physics, University of California, Berkeley, CA 94720, USA.
*These authors contributed equally to this work. **†Corresponding author. E-mail: martin.schultze@mpq.mpg.de (M.S.); dneumark@berkeley.edu (D.M.N.); srl@berkeley.edu (S.R.L.)**

Downloaded from www.sciencemag.org on December 14, 2014

materials (SM), section S2], and span a spectrum of extreme ultraviolet (XUV) photon energies covering 80 to 125 eV. This spectral range probes the density of empty electronic states in a time-resolved variant of x-ray absorption near-edge structure spectroscopy (XANES) (*16*). Figure 1B shows the recorded silicon *L*-edge absorbance A of a 250-nm-thick single crystalline free-standing silicon membrane (*17*) in the <100> orientation, as well as the derivative of the static absorbance $\partial A/\partial E$ as a function of the XUV photon energy E_{XUV} computed from the raw data, which facilitates assessment of excitation-induced broadening and shifts of the CB features. The above-edge structure of $A(E_{XUV})$ recorded by attosecond pulses is a sensitive probe of an extended region of the CB density of states, which is validated by calculated core-level transition spectra (see SM, section S9). Whereas the initial state of the electron undergoing the XUV probe transition is a $2p$ core orbital, the final state of both the XUV probe and the NIR pump excitation lies within the CB manifold. Electron interactions in the CB that alter the XUV transition thus render attosecond XUV absorption sensitive to the evolution of the pump-laser–induced CB population.

The silicon sample is irradiated by ultrashort NIR laser pulses comprising only a few oscillations of the electric field and attaining intensities between 10^{10} and 10^{14} W/cm^2 at the focus, corresponding to a vacuum peak amplitude E_0 of the electric field between 0.02 and 2 V/Å. The laser-induced transfer of electrons to the CB results in a wealth of modifications to the XUV transition. Figure 2A displays the derivative of the *L*-edge absorbance spectrum, $\partial A/\partial E$, as a function of time delay Δt between the NIR-pump and the XUV-probe pulses. Excitation of silicon by a few-cycle NIR laser pulse, with an intensity inside the sample of $I_{int} = 1.9 \times 10^{12}$ W/cm^2 (see SM, section S5 for the relation between I_{int} and the vacuum intensity I_{vac}), results in a global broadening of the substructures. Most important, as seen in Fig. 2B, the spectrum recorded with carrier-envelope-phase (CEP) stabilized NIR pump pulses evolves with a step-like behavior synchronized with the half-cycle period of the pump electric field oscillations, indicative of subfemtosecond population transfer and carrier-carrier scattering time scales. Further, an effective reduction of the measured band gap (band-gap narrowing) after the excitation is observed (Fig. 2C, blue line), and during the excitation pulse, a transient field–induced blue shift of the *L*-edge onset is recorded (Fig. 2C, red line). Finally, Fig. 2D shows a clear biexponential decay of the signal recorded over an extended delay range with time constants of 5 fs and 60 fs, respectively, indicating the transition between purely electronic initial response after the laser pulse and convoluted electronic-lattice dynamics at longer time scales.

The *L*-edge substructures exhibit a global broadening, as shown in Fig. 2A. Electron-electron interactions in the CB cause a modification of the dipolar oscillations associated with optical transitions in the time domain and thus have been previously indicated as the origin of spectral broadening in static studies (*18*). Similarly, the interaction of XUV probe-promoted electrons with the high density of light-field–injected carriers leads to rapid damping or dephasing of the oscillating XUV transition dipole observed as frequency domain broadening of the *L*-edge features during and after excitation.

Figure 2B shows the appearance of sharp steps in the *L*-edge transmittance evaluated at 100.35 eV, in synchrony with the CEP stabilized pump laser electric field oscillations. As the delay between the attosecond probe pulse and the few-cycle pump pulse is scanned, the steps on the rising edge of the time trace are formed when each half-cycle of the electric field of the pump pulse excites electrons from the VB to the CB, leading to synchronized observable changes to the *L*-edge transition. The rise time of each step

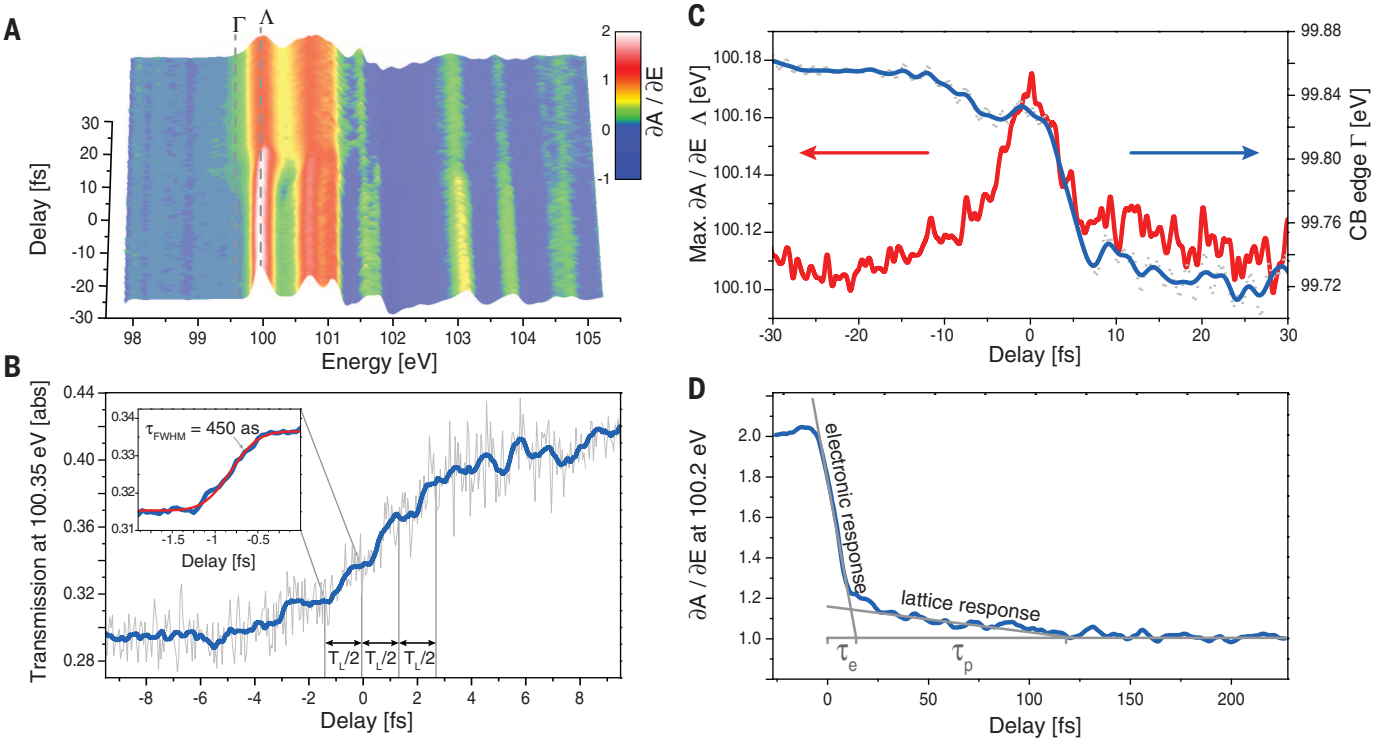


Fig. 2. Attosecond transient absorption spectroscopy of silicon. (**A**) The injection of electrons into conduction band states of silicon by a few-cycle near-infrared laser pump-pulse modifies the derivative of the XUV-absorbance $\partial A/\partial E$ plotted as a function of probe-photon energy E_{XUV} and time delay Δt between pump and probe pulse. The color scale represents the value of $\partial A/\partial E$. (**B**) A close-up of the temporal evolution of the XUV-transmission T (= 10^{-A}) at 100.35 eV reveals the increase of the signal amplitude in sharp steps synchronized with the laser electric field oscillations (also see Fig. 3B). The inset shows a fit to evaluate the step rise time. The blue line is a rolling average of the raw signal depicted in gray. (**C**) The position in energy of the first peak of the derivative [marked Λ in (A)] is plotted (red), along with the energy of the *L*-edge onset (marked Γ) evaluated at $\partial A/\partial E = 0.2$ (blue). (**D**) tracks the amplitude of $\partial A/\partial E$ at 100.2 eV, the position of the maximum $\partial A/\partial E$ before excitation. The straight lines indicate the different time scales over which electronic (τ_e) and nuclear (τ_p) dynamics occur.

in the subcycle evolution allows an estimate of the response time of the electronic ensemble by fitting an error function with a best-fit full-width half-maximum of ~450 attoseconds (see Fig. 2B inset and SM, section S4), providing an upper limit for the carrier-carrier scattering time in heavily photodoped silicon. An assumed longer response time would obscure the steplike signal evolution, as detailed in the SM, section S4. The findings substantiate the reported sub-1-fs (*19*) estimate for CB electron scattering times in dense electron-hole plasmas.

To develop an understanding of the time evolution of the physical processes dictating the interaction of the few-cycle laser pulses with silicon, we performed ab initio TD-DFT calculations (*20*) within the adiabatic local density approximation (*21*). By predicting the excitation process from first principles without tunable parameters, the simulations quantify the increase in the number of excited electrons with time and their distribution in real and momentum space after the excitation. For intensities up to one order of magnitude of the limit for dielectric breakdown of the sample, the TD-DFT treatment of the excitation dynamics predicts an increase in the number of excited electron hole pairs (Fig. 3A) following the oscillations of the laser electric field, analogous to the stepwise signal evolution shown in Fig. 2B. As input for the simulations, we employed the temporal electric field evolution of the few-cycle NIR driver pulses (Fig. 3B), as recorded by means of attosecond streaking (*22*) of photoelectrons released from neon (see SM, section S2).

When the NIR excitation field is acting on the system, it mixes the CB and VB states and causes tunneling of electrons across the direct band gap. Remarkably, the theory predicts a deviation of the excitation energy per electron-hole pair (Fig. 3C) and of the number of excited electrons per atom (Fig. 3D), from the multiphoton to the field-induced tunneling regime. Field-induced tunneling dominates for laser intensities that correspond to a value of the Keldysh adiabaticity parameter $\gamma < 1$, defined as $\gamma = \frac{\omega}{eE_0}\sqrt{m^*\Delta_g}$, where Δ_g, ω, and m^* are the band gap of the material, the laser frequency, and the reduced mass of the electron hole pair (here, 0.25 m_e), respectively (*23*). At the intensity of 1.9 × 10^{12} W/cm^2 used in the experiment, $\gamma \sim 0.5$, which suggests that the excitation mechanism is within the tunneling regime.

The insets in Fig. 3 display a spatial representation of the ultrafast band-gap excitation derived from the TD-DFT calculation. Modifications of the spatial electronic density distribution along a cut through the eight-atom unit cell in the [011] × [100] plane are shown relative to the ground state. During the laser pulse (panels t_1 and t_2), field-driven dipolar oscillations along the silicon-silicon bond dominate. After the laser electric field passes (panel t_3), some electronic density along the interatomic bonds is depleted, and the density is relocated to the interstitial regions that represent the mobile antibonding conduction band states of the crystal.

The validity of the predicted CB electron energy distribution and the identification of field-induced tunneling as the main excitation mechanism are corroborated by the experimental observations at a laser intensity of I_{int} = 1.9 × 10^{12} W/cm^2. The calculations indicate that the occupation number profile of hole states exhibits a narrow distribution below the VB (mean value 0.5 eV below VB maximum) compared with the electrons in the CB, where the mean value of the distribution is at 2.1 eV above the CB minimum. Carrier-induced changes to the XUV absorbance are detected experimentally up to 6 eV above the band gap (see Fig. 2A and SM, section S6), beyond which the signal-to-noise ratio limits further measurements.

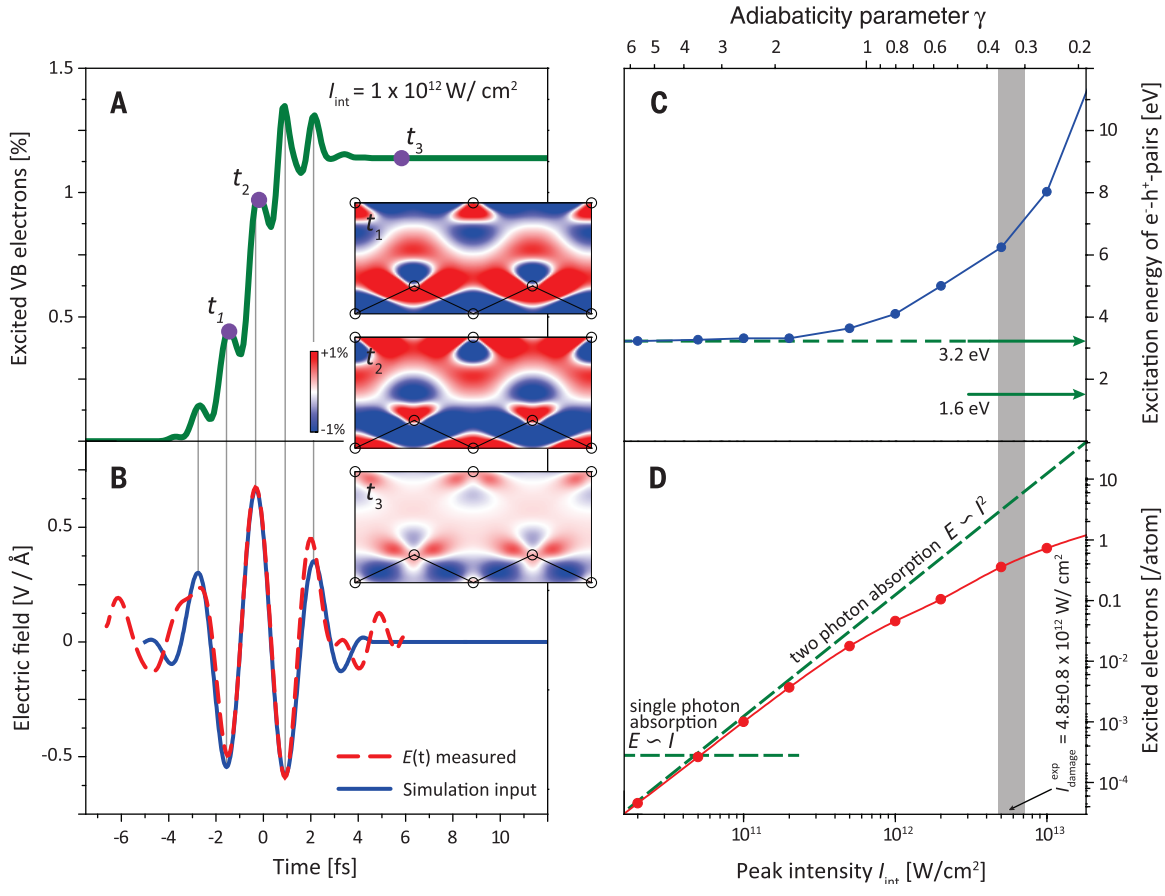


Fig. 3. Theoretical simulations of sub-femtosecond population transfer in silicon. (**A**) Number of electrons excited into the conduction band by a 4.8-fs laser pulse of peak intensity inside silicon of 1 × 10^{12} W/cm^2, as predicted by TD-DFT calculations. The extent of population transfer is quantified as a percentage of the initially available valence band electrons. As input for the calculations, a theoretical laser waveform [blue line in (**B**)] was used that was fitted to the laser electric field evolution extracted from an attosecond streaking measurement [red dashed line in (B)]. Gray lines between (A) and (B) identify the laser field extrema and the corresponding oscillations in (A). The insets show the calculated electronic density distribution during (t_1 and t_2) and after (t_3) the passage of the excitation pulse along the crystal plane defined by the [011] and [100] axes. The red and blue colors represent enhancement and depletion of electronic density, respectively, relative to the ground-state distribution. The calculated (**C**) mean excitation energy per electron-hole pair and (**D**) number of excited electrons plotted as a function of laser peak intensity inside the material deviate from the expectations for two-photon absorption behavior for values of the adiabaticity parameter <1, indicating transition into the light-field–induced tunneling regime.

These observations are irreconcilable with a multiphoton excitation mechanism, which would promote electrons to states concentrated at the band edge. In contrast, the electrons have an increasing amount of excess energy above the band gap as the laser intensity is increased, in agreement with experimental observations of an intensity-dependent extent of above-gap changes in the XUV transitions (see SM, section S6). The width of the above-gap excitation distribution is in accord with TD-DFT calculations, which predict that the number of excitation events compared with the expectation based on 2-photon absorption is truncated (Fig. 3, C and D). Concomitantly, the theory allows us to explore the intensity dependence of the population transfer and suggests that below the experimentally observed threshold for permanent sample damage of I_{int} = 4.8×10^{12} W/cm^2 (I_{vac} = 2.7×10^{13} W/cm^2) (see SM, section S5), more than 10% population transfer can be established during the few-cycle laser pulse (Fig. 3D).

After the NIR excitation, the measurements reveal a lasting shift of the *L*-edge absorption onset (i.e., the CB minimum) to a 121 meV lower value, as evaluated at $\partial A/\partial E$ = 0.2 of the *L*-edge spectral derivative (blue line in Fig. 2C). We performed core-level absorption calculations (*24*) with the electronic excitation density distribution obtained from the TD-DFT treatment (details in the SM, section S7) to obtain transient core-level spectra and relate the observed shifts to the number of excited carriers. The calculations reveal that Pauli state–blocking under these conditions has an insignificant influence on the XUV transition, owing to the dilute above-gap state occupation resulting from the tunneling process. The calculations furthermore estimate a core-level shift ΔE_{CL} following the band-gap excitation of 43 meV toward lower binding energy (see SM, section S9), rooted in the additional screening response of the excited carriers. This prediction on the single-particle level of theory disregards electron-electron interactions in the CB and reflects only 30% of the experimentally observed lasting red shift of the CB onset. In contrast, many-body core absorption theory, so far restricted to an equilibrated electronic CB population, confirms that the observed effective red shift of the CB onset (78 meV) is due to electron-electron interactions causing band-gap narrowing and increased above-gap electron scattering rates (see SM, section S9).

Valence and conduction bands are reported to shift by identical amounts in the presence of electron-hole pairs (*18*), and comparison with band-gap displacements observed as a function of doping concentration (*12*) suggests that ~10^{21}/cm^3 electron-hole pairs are created by the pump pulse. In striking agreement with this estimate, the ab initio TD-DFT calculations predict that after a 4.8-fs NIR pulse with an intensity of $I_{int} = 1 \times 10^{12}$ W/cm^2, 1.2% of all the valence electrons are excited into the CB, resulting in a carrier density of ~2.3×10^{21} cm^{-3} and a corresponding increase in conductivity of seven orders of magnitude (*25*) (see SM, section S12). The excitation-induced broadening of the spectral features (Fig. 2A) is observed for at least 200 fs, suggesting the persistence of the population transfer beyond the time scale sampled in these experiments.

Transiently, the experimental spectrum manifests a field-induced blue shift of the *L*-edge transition during the pump pulse (Fig. 2C, red line). In the presence of the excitation field, the XUV transition has to bridge the core-level-to-CB energy difference and provide the ponderomotive energy for the electron to oscillate in the field (*26*), analogous to the core-to-continuum transition in atoms dressed by an external field (*27*). The observed transient blue shift of ~80 meV is smaller than the calculated ponderomotive energy of 112 meV for a laser intensity of I_{int} = 1.9×10^{12} W/cm^2, owing to the competing transient red shift of the XUV transition due to the quadratic Stark effect. This acts to lower both the 2*p* core level and the CB minimum, but by different magnitudes due to the higher polarizability of the delocalized CB states. Core-level absorption calculations indeed predict a counteracting red shift of the *L*-edge transition under the influence of the laser field but do not yield quantitative agreement with the experiment. This discrepancy likely reflects disregard of the dynamical polarizability of the partially populated CB manifold, a subject for future theoretical developments. Both effects appear fully reversible and disappear after the passage of the few-cycle pump pulse.

On the time scale characteristic of lattice dynamics, tens of fs and longer, the experiments reveal participation of lattice modifications. Figure 2D shows the amplitude of $\partial A/\partial E$ at 100.2 eV, the photon energy corresponding to maximum $\partial A/\partial E$ before excitation. About 80% of the persistent pump-induced changes of the XUV absorbance are established immediately during the laser electric field and can be attributed to the ultrafast electronic response already discussed. When the temporal behavior of the amplitude of the derivative is fit with a biexponential decay (see SM, section S10) with time constants of τ_e and τ_p, the residual ~20% change in the amplitude is governed by a time scale of $\tau_p = 60 \pm 10$ fs, characteristic of the fastest optical phonon, which has a period of 64 fs (*14*, *15*). The TD-DFT calculations establish that this periodic modulation of the interatomic distances is set in motion when electrons tunnel from VB states located in the interatomic bond region to the interstitial CB states (Fig. 3, insets). As a result, the XUV transition exhibits an inhomogeneous broadening as the transmitted XUV pulse averages over many periods of the phonon motion and the associated interatomic distances. Indeed, the core-level absorption calculations indicate that an additional 20% spectral broadening of the XUV transition beyond the electronic response is expected over one full period of the longitudinal optical (LO) phonon vibration assuming 6-picometer atomic displacement (SM, section S10). In these experiments, we do not expect to observe periodic recurrences of the LO phonon in the time span sampled because of the insensitivity of the transmitted XUV pulse to the phase of the lattice vibration. Furthermore, fully coherent motion sets in only beyond 2 ps (*15*), before which higher harmonics of the vibration interfere and result in aperiodic modulation of the signal. The observation of a temporal hierarchy in the coupled electron-lattice response validates the widely employed phenomenological two-temperature model describing energy relaxation in solids (*28*). There, an impulsive excitation is assumed to primarily raise the electronic temperature followed by a heating of the lattice by the hot electrons. These results also highlight the fact that, due to the mismatch between almost instantaneous electronic and delayed lattice response, ultrashort excitation pulses can cause percent-level carrier excitation quasi-instantaneously without substantial lattice distortion or sample damage.

The presented methodology could generally explore the dynamics of light-induced excitation processes in semiconductors independent of the excitation mechanism. The ability to separate electronic and lattice response to an initial population transfer will permit the investigation of postexcitation carrier interactions and relaxation dynamics.

Through analysis of the experimental *L*-edge spectra at different time delays, in combination with TD-DFT and core-level absorption calculations, these studies reveal the time scale of carrier-carrier interactions and identify their role in the band-gap reduction of semiconductors. The attosecond temporal resolution allows identification of light-field–driven electron tunneling as the main process for band-gap excitation by few-cycle near-infrared laser pulses.

REFERENCES AND NOTES

1. J. Shah, *Ultrafast Spectroscopy of Semiconductors and Semiconductor Nanostructures* (Springer, Berlin, New York, 1996).
2. F. Krausz, M. I. Stockman, *Nat. Photonics* **8**, 205–213 (2014).
3. W. F. Koehl, B. B. Buckley, F. J. Heremans, G. Calusine, D. D. Awschalom, *Nature* **479**, 84–87 (2011).
4. P. T. Greenland *et al*., *Nature* **465**, 1057–1061 (2010).
5. M. Lenzner *et al*., *Phys. Rev. Lett.* **80**, 4076–4079 (1998).
6. D. Du, X. Liu, G. Korn, J. Squier, G. Mourou, *Appl. Phys. Lett.* **64**, 3071 (1994).
7. M. Gertsvolf, M. Spanner, D. M. Rayner, P. B. Corkum, *J. Phys. At. Mol. Opt. Phys.* **43**, 131002 (2010).
8. P. C. Becker *et al*., *Phys. Rev. Lett.* **61**, 1647–1649 (1988).
9. W. Kuehn *et al*., *Phys. Rev. Lett.* **104**, 146602 (2010).
10. J. Noffsinger, E. Kioupakis, C. G. Van de Walle, S. G. Louie, M. L. Cohen, *Phys. Rev. Lett.* **108**, 167402 (2012).
11. M. Beye *et al*., *New J. Phys.* **12**, 043011 (2010).
12. R. J. Van Overstraeten, R. P. Mertens, *Solid-State Electron.* **30**, 1077–1087 (1987).
13. M. Schultze *et al*., *Nature* **493**, 75–78 (2013).
14. Y. Shinohara *et al*., *Phys. Rev. B* **82**, 155110 (2010).
15. M. Hase, M. Katsuragawa, A. M. Constantinescu, H. Petek, *Nat. Photonics* **6**, 243–247 (2012).
16. D. C. Koningsberger, R. Prins, Eds., *X-ray Absorption: Principles, Applications, Techniques of EXAFS, SEXAFS, and XANES* (Wiley-Interscience, New York, ed. 1, 1988).
17. Silicon crystals 250 nm thick, lightly p-doped (Boron dopant, 1.5×10^{15}/cm^3), have been obtained from Norcada, Inc. For the subcycle-resolved measurement, the samples have been thinned in a plasma etching protocol to a final layer thickness of 140 nm.
18. L. Viña, M. Cardona, *Phys. Rev. B* **29**, 6739–6751 (1984).

19. D. Hulin, A. Mourchid, R. Vanderhaghen, P. M. Fauchet, in *Ultrafast Phenomena VII SE - 85*, Springer Series in Chemical Physics. C. Harris, E. Ippen, G. Mourou, A. Zewail, Eds. (Springer, Berlin Heidelberg, 1990), vol. 53, pp. 282–284.
20. K. Yabana, T. Sugiyama, Y. Shinohara, T. Otobe, G. F. Bertsch, *Phys. Rev. B* **85**, 045134 (2012).
21. J. P. Perdew, A. Zunger, *Phys. Rev. B* **23**, 5048–5079 (1981).
22. R. Kienberger *et al*., *Nature* **427**, 817–821 (2004).
23. L. V. Keldysh, *Sov. Phys. J. Exp. Theor. Phys.* **20**, 1307–1314 (1965).
24. D. Prendergast, G. Galli, *Phys. Rev. Lett.* **96**, 215502 (2006).
25. W. R. Thurber, R. L. Mattis, Y. M. Liu, J. J. Filliben, *J. Electrochem. Soc.* **127**, 2291–2294 (1980).
26. A. Srivastava, R. Srivastava, J. Wang, J. Kono, *Phys. Rev. Lett.* **93**, 157401 (2004).
27. P. Radcliffe *et al*., *New J. Phys.* **14**, 043008 (2012).
28. P. B. Allen, *Phys. Rev. Lett.* **59**, 1460–1463 (1987).

ACKNOWLEDGMENTS

We acknowledge fruitful discussions with P. Feulner and help with sample preparation from R. Rivers. The experimental work is supported by the Office of Assistant Secretary of Defense for Research and Engineering through a National Security Science and Engineering Faculty Fellowship (NSSEFF), with additional funding from the Defense Advanced Research Projects Agency PULSE program through grant W31P4Q-13-1-0017. The W. M. Keck Foundation and the Department of Energy under contract DE-AC03-76SF00098 are acknowledged for additional experimental equipment. S.R.L. and D.M.N. acknowledge Multidisciplinary University Research Initiatives from the Army Research Office (WN911NF-14-1-0383) and the Air Force Office of Scientific Research. S.A.S and K.Y. performed computations at the Institute of Solid State Physics, University of Tokyo under Japan Society for the Promotion of Science KAKENHI grants 23340113 and 25104702. M.S. was supported by a Marie Curie International Outgoing Fellowship (FP7-PEOPLE-2011-IOF). C.D.P. and D.P. performed work at the Molecular Foundry, supported by the Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under contract DE-AC02-05CH11231, with computing resources at Lawrence Berkeley National Laboratory and the National Energy Research Scientific Computing Center (NERSC).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/346/6215/1348/suppl/DC1
Materials and Methods
Figs. S1 to S16
References (*29*–*59*)

22 August 2014; accepted 10 November 2014
10.1126/science.1260311

NANOLITHOGRAPHY

Large-scale nanoshaping of ultrasmooth 3D crystalline metallic structures

Huang Gao,[1,3]* Yaowu Hu,[1,3]* Yi Xuan,[2,3]* Ji Li,[1,3] Yingling Yang,[1,3] Ramses V. Martinez,[4,5] Chunyu Li,[3,6] Jian Luo,[7] Minghao Qi,[2,3] Gary J. Cheng[1,3,8]†

We report a low-cost, high-throughput benchtop method that enables thin layers of metal to be shaped with nanoscale precision by generating ultrahigh-strain-rate deformations. Laser shock imprinting can create three-dimensional crystalline metallic structures as small as 10 nanometers with ultrasmooth surfaces at ambient conditions. This technique enables the successful fabrications of large-area, uniform nanopatterns with aspect ratios as high as 5 for plasmonic and sensing applications, as well as mechanically strengthened nanostructures and metal-graphene hybrid nanodevices.

Nanoscale metallic structures and their possible uses are under investigation in a variety of fields such as plasmonics (*1*), electronics (*2*), and biosciences (*3*). However, the large-scale manufacture of such structures with high fidelity and quality [e.g., ultrasmooth surfaces, sharp corners, three-dimensional (3D) shapes, and high crystallinity] represents a substantial challenge. Although nanoimprint lithography is a useful method of fabricating nanometer-scale patterns on polymers (*4*, *5*) and has been adapted to metallic glasses (*6*), direct nanoimprinting of crystalline metals is still generally infeasible because of the limitations on formability arising from (i) fluctuations of plasticity at the nanoscale caused by localized dislocation bursts (*7*, *8*), (ii) size effects in plasticity (*9*, *10*), and (iii) grain size effects, which generally limit the feature size to be larger than the grains (*11*, *12*). These limitations can be circumvented by using nanocrystalline or amorphous metals (*6*), heating the sample close to melting temperature (*13*), or using a superhard mold (*14*). But these methods either place limits on the materials that can be used, or operate at high temperatures with serious drawbacks [e.g., the heating and solidification cycle in hot embossing (*14*) or melting (*15*) leads to high surface roughness, especially in sub–100-nm metallic structures]. Moreover, the low crystallinity in metallic glasses limits their functionality because of electric and magnetic losses (*16*). Similarly, fabrication of metal nanopatterns by direct nanoimprinting of nanoparticles (*17*) incurs size effects of the particles during their deposition into the nanomolds and is therefore not a good choice for ultrafine patterns. Currently, fabrication of metallic nanopatterns relies on time-consuming, multistep approaches consisting of electron beam lithography (*18*, *19*) or template stripping (*20*, *21*). However, high resolution, high crystallinity, and low sidewall roughness cannot be achieved simultaneously, even with state-of-the-art equipment.

[1]School of Industrial Engineering, Purdue University, West Lafayette, IN 47907, USA [2]School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907, USA [3]Birck Nanotechnology Center, Purdue University, West Lafayette, IN 47907, USA [4]Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA [5]Madrid Institute for Advanced Studies, IMDEA Nanoscience, Ciudad Universitaria de Cantoblanco, 28049 Madrid, Spain. [6]School of Materials Engineering, Purdue University, West Lafayette, IN 47907, USA. [7]Department of NanoEngineering, University of California, San Diego, La Jolla, CA 92093, USA. [8]School of Mechanical Engineering, Purdue University, West Lafayette, IN 47907, USA.
*These authors contributed equally to this work. **†Corresponding author. E-mail: gjcheng@purdue.edu**

Figure 1A summarizes the most common direct shaping processes, materials, ranges of their respective processing strain rates, and processing temperatures. A key barrier that impedes the fabrication of smooth ultrafine crystalline metallic 3D nanostructures is the high strain rate required to activate superplasticity in metals, due to the need of ultrafine grain/particle sizes to generate superplastic flows (*22*, *23*). Unfortunately, the existing imprinting methods cannot provide sufficient high strain rates to generate 3D metallic crystalline nanostructures.

We demonstrate "laser shock imprinting" (LSI), a cost-effective direct nanoshaping method for high-throughput fabrication of smooth 3D crystalline nanostructures at ambient conditions. LSI uses a laser shock to compress metallic sheets into a silicon nanomold with a variety of shapes. We have used the nanomolds for over a hundred times with no damage or degradation in performance (fig. S1). This technique enables large-scale direct fabrication of smooth ultrafine metallic nanopatterns as small as 10 nm for mass production (fig. S2). LSI uses ultrahigh–strain rate deformation at room temperature to overcome the limitations in nanoscale formability of coarse-grained metals. Additionally, LSI can be applied to a broad range of metals, including metals as hard as Ti (fig. S3).

Figure 1B schematically illustrates the LSI process. We used a Nd:YAG laser pulse (0.3 to 1.4 GW/cm^2, wavelength 1064 nm, pulse duration ~5 ns) to irradiate an ablation coating layer (10 μm graphite) in direct contact with the sample. The sublimation of the ablation layer by the laser pulse generates a shock wave that, constrained by the confinement layer (glass or water), propagates through the metallic sheet and plastically deform it conformably on the underlying silicon mold with nanoscale resolution. According to Fabbro's model (*24*), the peak pressure of the shock wave generated by the laser pulse reaches 0.85 to 1.83 GPa, enabling the generation of strain rates on the order of 10^6 to 10^7 s^{-1} during the laser shock, depending on the applied laser intensity and pulse duration.

We fabricated silicon nanomolds by e-beam lithography (EBL) and reactive ion etching (RIE) or wet etching. Atomic layer deposition (ALD) was used to deposit an ultrathin Al_2O_3 layer (5 to

Downloaded from www.sciencemag.org on December 14, 2014

10 nm) to both increase the abrasive resistance of the nanomold and to help achieve the fabrication of sub–20-nm features in the Si wafers. We deposited an ultrathin layer (1 to 5 nm) of polyvinylpyrrolidone (PVP) or Au thin film (5 to 10 nm) on the surface of metal to act as a dry lubricant to reduce the friction during LSI processing and to improve the lateral resolution (fig. S4). Nonlubricant forming could also be used for ultrafine patterns (<30 nm) to ensure the precision.

Figure 2 shows several unique examples of metal nanostructures fabricated with LSI. Pyramidal nanostructures have been commonly used in surface-enhanced Raman scattering (SERS), due to the strong field enhancement achieved when surface plasmons propagating along nonadjacent faces converge at the tip of the pyramid (*21*). Most of the pyramidal nanoarrays fabricated previously were not optimal in achieving a strong plasmonic enhancement, due to the large base size of the pyramids (5 to 20 μm), the amorphous structure of the plasmonic material, and the limited sharpness of the tip due to the metal deposition process. Figure 2A shows a large-area LSI of nanopyramids on the surface of a 2.5-μm-thick sheet of polycrystalline Ag (with micrometer-scale grain size). The smooth surfaces and sharp tips of the pyramids indicate the outstanding formability induced by LSI. Pyramids with radius of curvature below 5 nm at their apex have been fabricated using a laser power of 1.4 GW/cm^2 (Fig. 2A, inset), and the possible mechanisms for forming such a sharp tip is discussed in supplementary materials. The sharpness of the nanotip greatly increases with the laser intensity (fig. S5).

The fabrication of nanoscale fishnet designs enabled the making of left-handed metamaterials with negative values of effective permittivity ϵ and permeability μ. Figure 2B shows a uniform fishnet structure (linewidth 30 nm, spacing 60 nm) fabricated on the surface of a polycrystalline Ag thin film by LSI. Currently, there is interest in the fabrication of hyperbolic metamaterials to realize useful functionalities in emerging metamaterials based on light (*25*). Unfortunately, the fabrication of many promising high-aspect-ratio periodic arrays of metallic nanostructures is difficult to achieve. LSI can fabricate smooth crystalline structures on Al thin foils with aspect ratios of ~3, as shown in Fig. 2C (width 220 nm, height 645 nm), using a laser intensity of 0.85 GW/cm^2. Higher aspect ratios could be achieved by increasing the laser intensity. Figure 2D shows an array of nanogears imprinted by LSI from a cold-rolled Al thin foil. The gear tooth sizes (length ~300 nm, height ~50 nm, width ~200 nm, radius at tooth bottom ~50 nm) are much smaller than the original grain sizes of the Al foil (~1.4 μm; see fig. S6A).

The triangular (V) groove-like nanostructures support the propagation of channel plasmon polaritons (*26*), which are attractive for applications in miniaturization of photonic circuits (*27*). Most V-grooves are fabricated by FIB, which is a very slow and low-volume prototyping process. With LSI and a suitable Si mold, we achieved a large-area array of nanoscale V-grooves with aspect ratios of 1:5 (width 100 nm, height 500 nm, V-angle 14°) and sharp side walls in Ag (Fig. 2E), with laser intensity of 0.334 GW/cm^2.

LSI is also a suitable technique to pattern hard metals, such as Ti with a melting temperature of 1668°C and a Vicker's hardness 5 times that of aluminum. The ultrasmooth nanostructures fabricated on the surface of a 2.5-μm-thick layer of Ti (Fig. 2F and fig. S3) demonstrate its good formability using LSI. We recognize that LSI, at its present level of development, still has some limitations, including the relatively high fabrication cost of the nanomolds (which, however, can be reused for hundreds of imprintings). It is also a challenge to transfer the imprinted metal nanostructures onto another substrate. LSI, however,

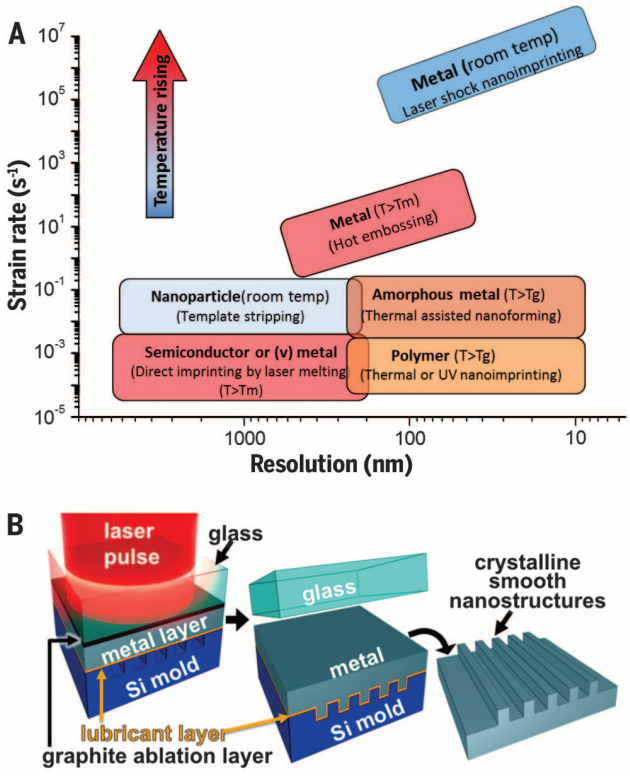


Fig. 1. Schematic illustration of the LSI setup and comparisons with other imprinting methods. (**A**) Comparison of the current imprinting techniques in terms of processing temperature, materials, resolution, and strain rate. (**B**) Summary of the LSI process.

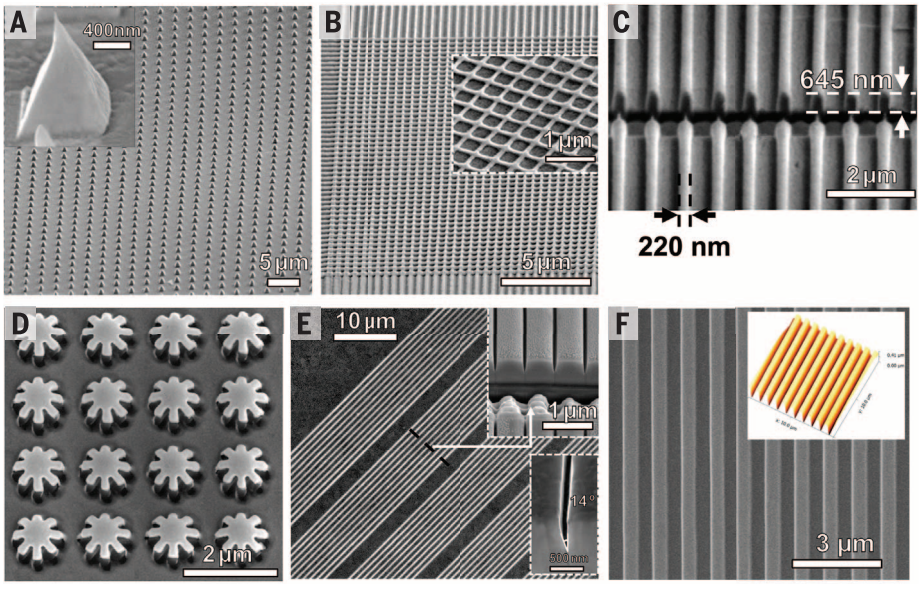


Fig. 2. SEM images of several complex 3D nanostructures fabricated by LSI. (**A**) Ag nanopyramids. (**B**) Ag fish-net (linewidth 30 nm, spacing 60 nm). (**C**) Nanobars with aspect ratio of ~3. (**D**) Array of nanogears imprinted on cold-rolled aluminum foil. (**E**) Array of triangular V-grooves (width 100 nm, depth 500 nm). (**F**) Array of nanotrenches (width 500 nm) in Ti thin film.

is versatile and scalable and, with an appropriate choice of materials for the nanomolds and selection of laser shock conditions, could enable high-throughput fabrication of sub–10-nm 3D nanostructures.

Morphological fluctuations during microscale deformation at low strain rates can result from the generation of localized dislocation avalanches (*9*, *10*). Our results, however, show that LSI, at high strain rates, enables the direct imprinting over large areas of uniform and smooth arrays of complex nanofeatures on the surface of crystalline metals. LSI overcomes both the limit of nanoscale formability of metallic materials and the size effect at the nanoscale.

To understand the underlying mechanisms of LSI, we investigated the microstructure of Al nanobars imprinted with LSI by combining high-resolution transmission electron microscopy (HRTEM) with molecular dynamics (MD) and finite element method (FEM) simulations. The methods and results are documented in supplementary materials, which are briefly summarized in Fig. 3 and the following text. The as-received Al sample has only one large grain through film thickness (fig. S6). After LSI, significant reductions in and nonuniform distribution of grain sizes can be observed. HRTEM images (Fig. 3B and fig. S6B) show that relatively large grains are present within and around the nanobars, while small grains with sizes of tens of nanometers are present in between the nanobars. The corresponding inverse fast Fourier transform (IFFT) images (Fig. 3C) reveal a low dislocation density inside the patterned nanostructures, while higher dislocation densities are found near the corners of the nanostructures. From MD simulation, strains around the edges and corners of the nanomold are much higher than at areas inside the nanomold (Fig. 3A). The relative low dislocation density within the patterned nanostructures can be explained from the nanoscale geometrical confinement effect on the propagation of dislocation (*10*) and, possibly, by a low- temperature dynamic recrystallization, which is often found at high strain rate deformation (*28*) and supported by our MD simulation (Fig. 3A) in that the strain field inside the nanomold is much lower than that outside of the nanomold. A local temperature rise inside the nanomolds, which is supported from FEM simulation (fig. S7), also facilitates dynamic recrystallization. In LSI, the nanosized grains formed at the initial stage of deformation are the keys to the nanoscale formability. This can be explained by the coalescence of dislocations and their reorganization into grain boundaries to reduce grain sizes during the imprinting process; subsequently, grain boundaries act as dislocation sources and sinks, which leads to the ultrasmooth nanoshaping. Grain boundary disordering, as indicated by MD simulation (fig. S9), can also serve as an effective sink to adsorb dislocations. This interaction between dislocations and grain boundaries hinders dislocation avalanches and strain burst, promoting a uniform deformation and maintaining smooth surfaces. Additionally, the large surface area of the nanopatterned structures can further absorb the dislocations during the LSI process. This hypothesized mechanism is consistent with our MD simulation (figs. S8 to S11) and a prior study (*9*) which suggested that grain boundaries hinder avalanche propagation and that the size of strain bursts is reduced when the grain size decreases.

Shown in Fig. 3, D and E, are the different microstructures of Al and Ag nanopyramids fabricated using the same Si nanomold (base width 800 nm) and laser intensity (0.87 GW/cm^2). For Al, a relative high stacking fault energy facilitates the cross-slip of partial dislocations, and the recrystallization temperature is low (175°C). As a result, Al nanopyramids are largely converted into single grains and low-angle grain boundaries are dominant in the microstructure (Fig. 3D). On the other hand, face-centered cubic (fcc) metals with low stacking fault energies that result in large separations between the partial dislocations such as Au and Ag, are difficult to have cross-slip and prone to work hardening. The dislocation density in these materials is increased by the successive generation of dislocations. Additionally, the high recrystallization temperature (970°C for Ag) impedes the efficient reorganization and annihilation of the dislocations. Figure 3E shows the microstructure of the Ag nanopyramids, in which low-angle grain boundaries and nanotwinning (with 60° misorientation) are dominant.

LSI not only improves the formability of materials but also the hardness of the fabricated nanostructures. Figure S12 shows the typical loading and unloading curves of as-received Al

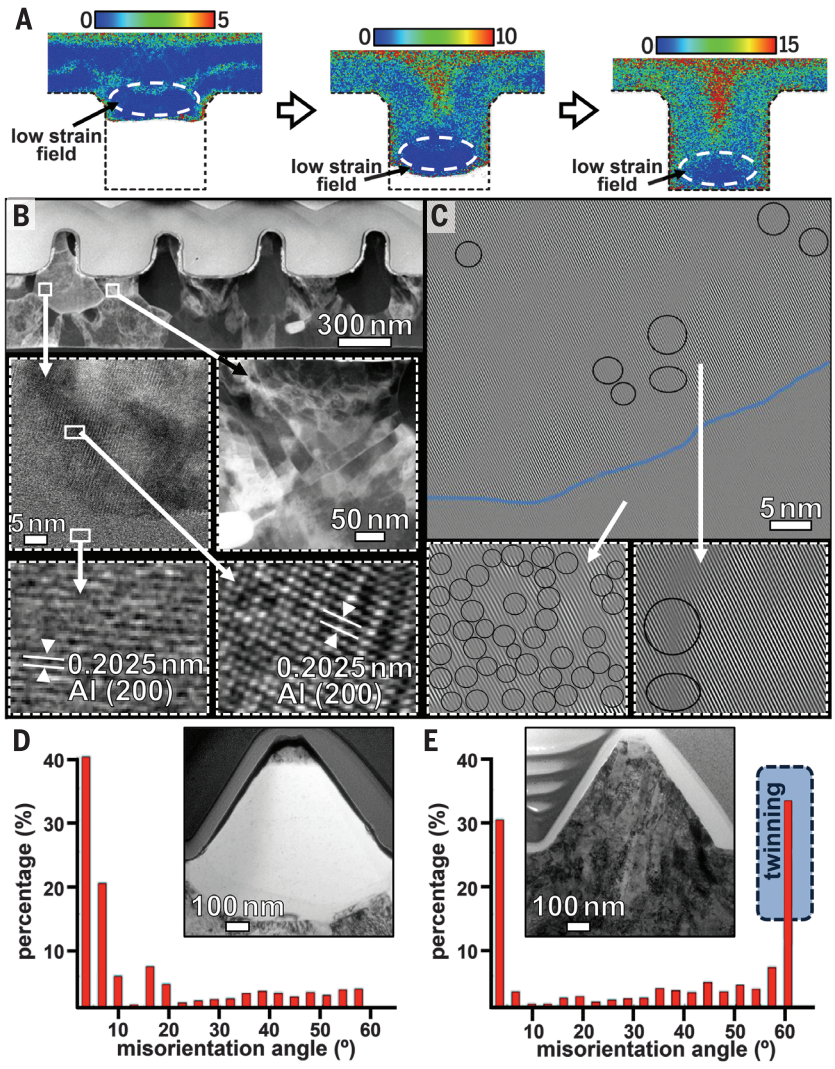


Fig. 3. Results of MD simulation and TEM characterization of imprinted specimens. (**A**) MD simulation results showing the strain distribution during LSI of a single crystal of Al (thickness 8 nm) to a nanomold (width and depth, 16.44 nm). (**B**) Cross-sectional STEM images of an Al thin film after LSI. Dark-field STEM image at high magnification showing the grain structure between the nanobars. HRTEM image of a large-angle grain boundary in the nanobar, showing the lattice structure of the crystals within and around the nanobars. (**C**) The corresponding IFFT image of the area within and around the nanobars. (**D** and **E**) TEM and grain boundary misorientation distribution of (D) an Al nanopyramid and (E) an Ag nanopyramid.

foil and the LSI processed Al foil during a nanoindentation cycle. The hardness of as-received samples is ~0.37 GPa, whereas the hardness at the nanobars fabricated by LSI is ~0.64 GPa. The improved strength of the nanostructures fabricated by LSI correlates with grain/subgrain refinement (fig. S13) and the number of residual dislocation structures (fig. S14).

Metamaterial perfect absorbers (MPAs) have attracted great attention for potential applications in emitter, sensing, detection, and wireless communication (*29*). Although MPAs operating at microwave and terahertz frequencies with micrometer structures could be easily achieved, fabricating nanometer features for visible and ultraviolet regions remains a challenge because of limited nanofabrication methods. Periodic aluminum nanotrenches with 10-, 20-, and 30-nm gaps were fabricated using the LSI method in a fast and inexpensive way via a single laser shot (laser intensity of 0.61 GW/cm^2) on a silicon nanomold (Fig. 4, B to D). We used atomic force microscopy (AFM) to confirm the ultrasmooth topography of the Al nanotrenches (fig. S2) and finite-difference time domain (FDTD) simulations to calculate the dependence of the absorbance over the gap width (Fig. 4E). The results revealed that the Al nanotrenches with 30-nm gaps have near-perfect absorption at λ = 436 nm (fig. S15 and fig. S16, A and B), thereby demonstrating MPA with a visible band. The surface-enhanced Raman scattering (SERS) of the Al nanotrenches fabricated by LSI depends on the matching of the incident and exciting radiation with plasmons on the substrate, which is controlled by the geometry of the fabricated structures. Figure 4F shows the application of Al nanopatterns fabricated by LSI as SERS substrate for graphene with an enhancement up to three orders of magnitude for the G and 2D peaks of graphene.

As a high–strain rate, cold-forming process, LSI is also suitable for the fabrication of hybrid nanostructures. Figure 4G illustrates the formation of graphene-metal hybrid nanostructures by direct imprinting a single layer of graphene grown on Cu foil and forming a graphene nanogap. Figure 4H shows the SEM image after deformation of graphene-Cu thin film into nanogaps by LSI with corresponding AFM image. The local field is probed using a near-field scanning optical microscope (NSOM) (Fig. 4I). As shown in the NSOM image, a high-intensity region has been detected at the center of the nanogap with laser excitation wavelength 633 nm. Because of the resolution limit of the NSOM tip, the two hot spots located at the structural corners are merged during the experiment, resulting in a single high-intensity slit. Graphene can be strained without breaking under certain conditions. The strainability of graphene by deforming to 3D nanoscale surface has been studied (*30*). Figure 4J shows the Raman spectrum of the graphene-Cu thin film after LSI under several laser conditions. When the laser

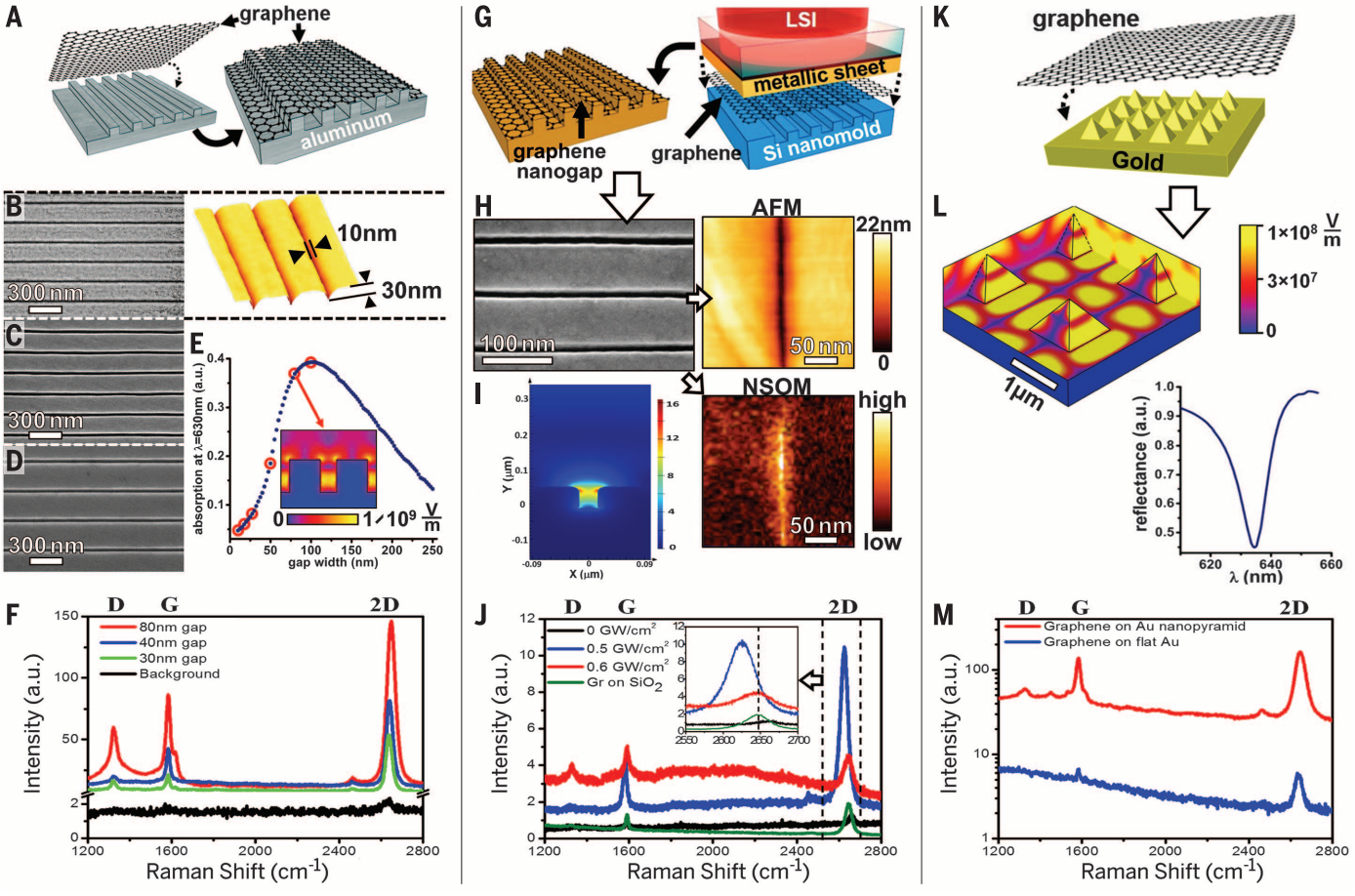


Fig. 4. Various ultrasmooth metallic and metal-graphene nanostructures made by LSI with enhanced electromagnetic and optical properties. (**A**) Schematic picture of transferring graphene on nanoshaped meal structures for SERS measurements. (**B** to **D**) SEM images of periodic aluminum trenches (10, 20, and 30 nm) after LSI, and an AFM image showing the ultrasmooth topography of the Al trench. (**E**) Absorption at λ = 630 nm of the Al nanotrenches fabricated by LSI vs. the gap width of the nanopatterns. (**F**) SERS measurement of nanopatterned Al with various gap width. (**G**) Illustration of the formation of graphene-Cu hybrid nanogaps, and transferring of graphene to Au nanopyramid arrays. (**H**) SEM and AFM image of a graphene-Cu nanogap after LSI (0.6 GW/cm^2). (**I**) Results on local field enhancement of the graphene-Cu nanogap by FDTD simulation and NSOM measurement. (**J**) Raman spectrum of graphene on SiO_2, graphene on Cu and nanopatterned graphene-Cu structures by LSI (0.5 GW/cm^2, 0.6 GW/cm^2). (**K**) Illustration of transferring graphene to Au nanopyramid arrays. (**L**) FDTD simulations results showing the field enhancement around the tip at the resonant frequency. (**M**) Raman spectra showing enhancement of Raman signatures of graphene on nanopyramid compared with that on flat surfaces.

intensity is larger than 0.6 GW/cm^2, graphene will be cut into nanogaps (Fig. 4H). The Raman spectrum shows that D peak (related to defect level) is high, while G peak is not shifted relative to that in the unstrained graphene, since strain is released after cut. When the laser intensity is less than 0.5 GW/cm^2, the graphene is continuous and conformal on nanoshaped Cu. This is indicated in the Raman spectrum (Fig. 4J) where the D peak is low and the G peak is shifted to the left because of straining.

In order to study the SERS of nanopyramids, graphene is transferred onto the Au nanopyramids array (Fig. 4K) to measure the localized plasmon-enhanced Raman spectra and verify the enhancement of the Raman signatures of graphene. As shown in Fig. 4M, the Raman signature of a monolayer of graphene (both G and 2D peak) is highly increased when it is placed on top of the nanopyramids, demonstrating the capability of local field enhancement of these structures. The ratio of 2D/G intensities has decreased from around 3 to 1, which could be a result from charge doping of graphene from the imprinted plasmonic nanostructures under laser excitation. FDTD simulations confirmed the field enhancement around the tip at the resonant frequency (Fig. 4L and fig. S16, C and D).

LSI is a high-throughput 3D nanoimprinting technique capable of producing nanoscale crystalline metallic nanostructures over a 6-inch wafer within 30 s, using laser pulses (pulse energies of 150 to 250 mJ, a beam size of 3 mm, and pulse frequency of 10 Hz). The crystalline nanostructures fabricated by LSI have several characteristics that make them especially suited for electronic, plasmonic, or sensing applications: (i) precise and ultrasmooth nanoshaping due to better formability and low-temperature processing; (ii) complex 3D nanoshaping, sub–20-nm lateral sizes, and aspect ratios up to 5 when fabricated using silicon nanomolds; (iii) beneficial crystalline metallic nanostructure for device performance, because it has the potential to improve electronic stability and to reduce optical loses, noise, and energy consumption. Our results show that LSI has the potential to provide new insights into the role of crystallinity in the electronic, optical, and mechanical behavior of metallic nanostructures. Further, LSI is an attractive fabrication method for the development of future electronics, optics, and sensing devices.

REFERENCES AND NOTES

1. L. Yin *et al.*, *Nano Lett.* **5**, 1399–1402 (2005).
2. E. Ozbay, *Science* **311**, 189–193 (2006).
3. J. R. Lakowicz, *Plasmonics* **1**, 5–33 (2006).
4. S. Y. Chou, P. R. Krauss, P. J. Renstrom, *Science* **272**, 85–87 (1996).
5. L. J. Guo, *Adv. Mater.* **19**, 495–513 (2007).
6. G. Kumar, H. X. Tang, J. Schroers, *Nature* **457**, 868–872 (2009).
7. M. D. Uchic, P. A. Shade, D. M. Dimiduk, *Annu. Rev. Mater. Res.* **39**, 361–386 (2009).
8. J. R. Greer, J. T. M. De Hosson, *Prog. Mater. Sci.* **56**, 654–724 (2011).
9. F. F. Csikor, C. Motz, D. Weygand, M. Zaiser, S. Zapperi, *Science* **318**, 251–254 (2007).
10. M. D. Uchic, D. M. Dimiduk, J. N. Florando, W. D. Nix, *Science* **305**, 986–989 (2004).
11. F. Vollertsen, H. Schulze Niehoff, Z. Hu, *Int. J. Mach. Tools Manuf.* **46**, 1172–1179 (2006).
12. W. Kurnia, M. Yoshino, *J. Micromech. Microeng.* **19**, 125028 (2009).
13. S. Y. Chou, C. Keimel, J. Gu, *Nature* **417**, 835–837 (2002).
14. K. A. Lister *et al.*, *J. Vac. Sci. Technol. B* **22**, 3257–3259 (2004).
15. B. Cui, C. Keimel, S. Y. Chou, *Nanotechnology* **21**, 045303 (2010).
16. R. F. Oulton, *Nat. Photonics* **6**, 219–221 (2012).
17. C. C. Liang *et al.*, *Opt. Express* **19**, 4768–4776 (2011).
18. N. Yu *et al.*, *Science* **334**, 333–337 (2011).
19. J. Lin *et al.*, *Science* **340**, 331–334 (2013).
20. W. Zhou, T. W. Odom, *Nat. Nanotechnol.* **6**, 423–427 (2011).
21. P. Nagpal, N. C. Lindquist, S. H. Oh, D. J. Norris, *Science* **325**, 594–597 (2009).
22. R. S. Mishra, T. R. Bieler, A. K. Mukherjee, *Acta Metall. Mater.* **43**, 877–891 (1995).
23. T. G. Nieh, J. Wadsworth, T. Imai, *Scr. Metall.* **26**, 703–708 (1992).
24. R. Fabbro, J. Fournier, P. Ballard, D. Devaux, J. Virmont, *J. Appl. Phys.* **68**, 775–784 (1990).
25. A. Poddubny, I. Iorsh, P. Belov, Y. Kivshar, *Nat. Photonics* **7**, 948–957 (2013).
26. S. I. Bozhevolnyi, V. S. Volkov, E. Devaux, T. W. Ebbesen, *Phys. Rev. Lett.* **95**, 046802 (2005).
27. D. K. Gramotnev, S. I. Bozhevolnyi, *Nat. Photonics* **4**, 83–91 (2010).
28. Q. Li, Y. Xu, Z. Lai, L. Shen, Y. Bai, *J. Mater. Sci. Technol.* **15**, 435 (1999).
29. C. M. Watts, X. Liu, W. J. Padilla, *Adv. Mater.* **24**, OP98–OP120 (2012).
30. J. Li, T. F. Chung, Y. P. Chen, G. J. Cheng, *Nano Lett.* **12**, 4577–4583 (2012).

ACKNOWLEDGMENTS

Supported by a National Research Council Senior Research Associateship (G.J.C.), the FP7 People program under the project Marie Curie IOF-275148 (R.V.M.), National Security Science and Engineering Faculty Fellowship N00014-15-1-0030 (J.L.); NSF grants CMMI-0547636, CMMI-1436305, CMMI-1120577, and CNS-1126688; NIH grant 1R01RR026273-01; Defense Threat Reduction Agency grant HDTRA1-10-1-0106; and Office of Naval Research grant N00014-11-1-0678. We thank S. Suslov for assistance with TEM and X. Xu for help with NSOM.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/346/6215/1352/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S16
References (*31*–*55*)

18 August 2014; accepted 14 November 2014
10.1126/science.1260139

MEMBRANES

Metal-organic framework nanosheets as building blocks for molecular sieving membranes

Yuan Peng,[1,2] Yanshuo Li,[1]* Yujie Ban,[1,2] Hua Jin,[1,2] Wenmei Jiao,[1,2] Xinlei Liu,[1] Weishen Yang[1]*

Layered metal-organic frameworks would be a diverse source of crystalline sheets with nanometer thickness for molecular sieving if they could be exfoliated, but there is a challenge in retaining the morphological and structural integrity. We report the preparation of 1-nanometer-thick sheets with large lateral area and high crystallinity from layered MOFs. They are used as building blocks for ultrathin molecular sieve membranes, which achieve hydrogen gas (H_2) permeance of up to several thousand gas permeation units (GPUs) with H_2/CO_2 selectivity greater than 200. We found an unusual proportional relationship between H_2 permeance and H_2 selectivity for the membranes, and achieved a simultaneous increase in both permeance and selectivity by suppressing lamellar stacking of the nanosheets.

Gas separation with membranes is an energy-efficient and environmentally friendly alternative to cryogenic and adsorptive or absorptive gas separation processes. Polymer membranes are subject to a trade-off between productivity (permeability) and efficiency (selectivity), known as Robeson's upper bound (*1*, *2*). Membranes based on molecular sieve materials are expected to overcome this limitation by relying on their ability to distinguish molecules based on size and shape (*3*–*8*). Molecular sieve nanosheets (MSNs) with large lateral area and small thickness are the most appropriate building blocks for ultrathin, and thus ultrapermeable, molecular sieve membranes (*9*). The permeance of such membranes is measured in gas permeation units (1 GPU = 10^{-6} cm^3 cm^{-2} s^{-1} cmHg^{-1} at STP). Tsapatsis *et al.* demonstrated the fabrication of molecular sieve membranes based on exfoliated zeolite nanosheets with thickness at the unit cell level (~3 nm) (*10*), whereas the types of zeolites that can be easily exfoliated are rather limited (*11*). Recently, graphene oxide (GO) nanosheets with selective defects were used to produce ultrathin membranes with thickness of as low as 1.8 nm (*12*). However, the measured H_2 permeance of these extremely thin GO membranes, ~300 GPUs, was still at the same level as conventional microporous membranes (*13*). This can be attributed to the low density of selective defects and their random distribution in the GO nanosheets.

[1]State Key Laboratory of Catalysis, Dalian Institute of Chemical Physics, Chinese Academy of Sciences, 457 Zhongshan Road, Dalian 116023, China. [2]University of Chinese Academy of Sciences, Beijing 100049, China.
*Corresponding author. E-mail: leeys@dicp.ac.cn (Y.L.); yangws@dicp.ac.cn (W.Y.)

Downloaded from www.sciencemag.org on December 14, 2014

In recent years, metal-organic frameworks (MOFs) have emerged as a family of nanoporous molecular sieves. A large number of layered MOFs have been reported (*14*), and they could serve as a diverse source of MSNs if they could be exfoliated to nanometer-scale thickness (*15*–*17*). However, structural deterioration and morphological damage have hampered success in obtaining high-quality MOF nanosheets, which has hindered the application of the nanosheets as building blocks for molecular sieve membranes. We report the preparation of 1-nm-thick MSNs with large lateral area and high crystallinity from layered MOF precursors and demonstrate their use in fabricating ultrapermeable membranes that have excellent molecular sieving properties for H_2/CO_2 separation. H_2/CO_2 membrane separation is considered one of the key technologies for zero-emission fossil fuel power generation and hydrogen production (*18*).

The layered MOF precursor exemplified here is poly[Zn_2(benzimidazole)$_4$], denoted $Zn_2(bim)_4$ (*19*). The corresponding scanning electron microscopy (SEM) images are shown in Fig. 1A. In this structure, two-dimensional (2D) layers are oriented normal to the *c* axis, connected by weak van der Waals interactions (Fig. 1A, inset). In the layers, each Zn ion is coordinated by four benzimidazole (bim) ligands in a distorted tetrahedral geometry, and each bim ligand bridges two Zn atoms via a bis-monodentate linkage (Fig. 1B). Powder x-ray diffraction (XRD) measurements confirm that our product is isostructural to a previously determined MOF (Fig. 1C) [ref. no. 675375, Cambridge Crystallographic Data Centre (CCDC)] (*20*). $Zn_2(bim)_4$ can also be obtained in high yield via hydrothermal transformation of the well-known three-dimensional zeolitic imidazolate framework ZIF-7 (*19*, *21*) (fig. S1), which implies that $Zn_2(bim)_4$ possesses excellent thermodynamic stability (fig. S2).

Conventional physical exfoliation may damage the in-plane structure of MOF nanosheets. Soft-chemical exfoliation is a potential alternative (*22*), but requires chemicals that may negatively affect application of the nanosheets. To address these problems, we developed a soft-physical process: Pristine $Zn_2(bim)_4$ crystals were first wet ball-milled at very low speed (60 rpm), followed by exfoliation in volatile solvent with the aid of ultrasonication. We found that a mixture of methanol and propanol is the most appropriate for this process (fig. S3), based on our analysis of a number of solvents. We hypothesize that wet ball-milling facilitates the penetration of small methanol molecules into the galleries of the layered $Zn_2(bim)_4$, and propanol helps to stabilize the exfoliated nanosheets by adsorbing on the sheets with its hydrophobic alkane tails. The colloid suspension of $Zn_2(bim)_4$ MSNs is clear but exhibits a Tyndall effect due to light scattering by the nanosheets in the colloid (Fig. 1D, inset).

Figure 1E illustrates the network structure of a single-layered $Zn_2(bim)_4$ MSN. The square $Zn_4(bim)_4$ unit, as shown in the red circle, can be considered to be the subunit of the 2D layer. The aperture size of the $Zn_4(bim)_4$ unit, a four-membered ring, is ~0.21 nm as estimated from crystallographic data (Fig. 1F). The effective pore size should be slightly larger, considering the structural flexibility of the sheet. Thus, we can expect that membranes based on $Zn_2(bim)_4$ MSNs will achieve a high selectivity of H_2 (0.29 nm) over CO_2 (0.33 nm) through molecular sieving. Moreover, unlike zeolite nanosheets, which consist of cages (or half cages) or zigzag channels (*10*), the pores of $Zn_2(bim)_4$ MSNs are constructed with four flat bim molecules, ensuring the rapid transport of H_2 molecules.

A colloidal dispersion containing ~15 mg/liter $Zn_2(bim)_4$ MSNs was obtained by removing the larger unexfoliated particles after being left to stand for 2 weeks. The colloid remained stable at room temperature for several months. $Zn_2(bim)_4$ MSNs with wrinkles were frequently observed under transmission electron microscopy (TEM), indicating the flexibility of the nanosheets (fig. S4), which is beneficial for achieving a conformal layer on macroporous supports with rough surfaces (*10*). An isolated $Zn_2(bim)_4$ MSN with side length of ~600 nm is shown in Fig. 2A, where folds and curled edges are observed. The nanosheet is very thin, as determined from the contrast in the TEM image. As is common for MOF materials characterized using high-resolution TEM (HRTEM), the $Zn_2(bim)_4$ nanosheets visibly and rapidly degraded during imaging, as a result of electron beam irradiation. Thus, it is difficult to visualize the lattice fringes of the (*hk*0) planes

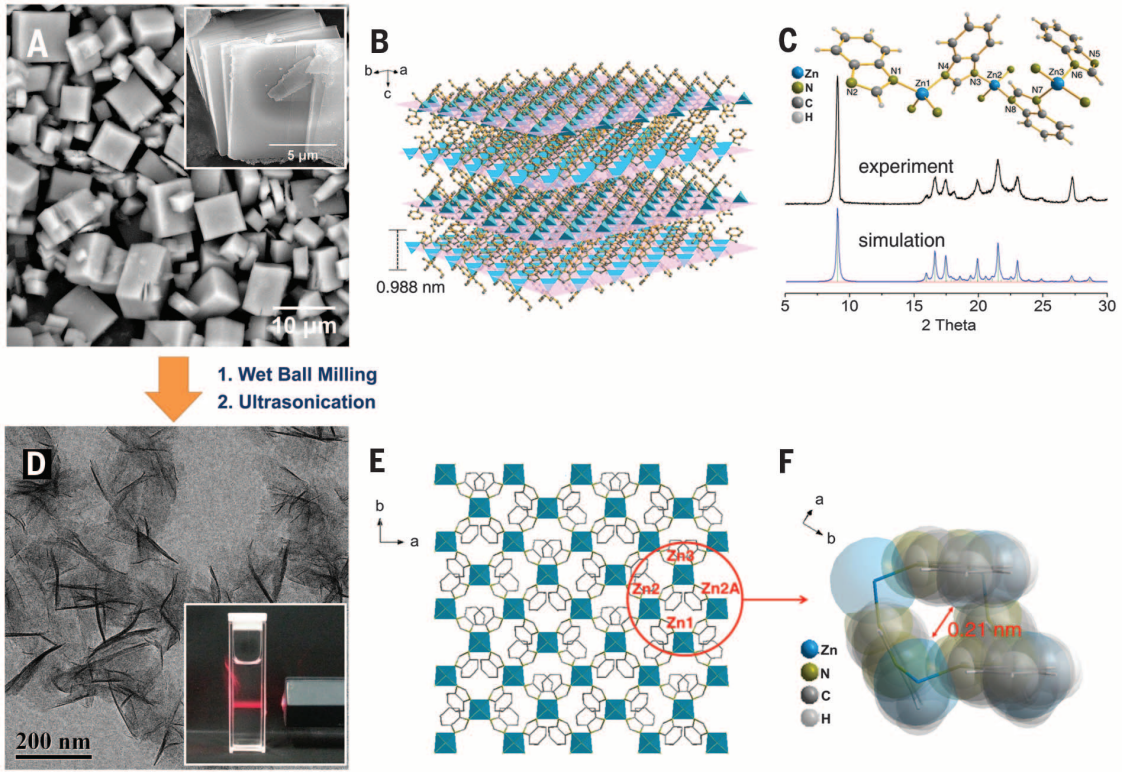


Fig. 1. Top-down fabrication of molecular sieve nanosheets. (**A**) Scanning electron microscopy (SEM) image of as-synthesized $Zn_2(bim)_4$ crystals. The inset image shows the typical flake-like morphology of $Zn_2(bim)_4$ crystals. (**B**) Architecture of the layered MOF precursor. The *ab* planes are highlighted in purple to better illustrate the layered structure. (**C**) Powder XRD patterns of $Zn_2(bim)_4$. The top trace is the experimental pattern, whereas the bottom trace is the pattern simulated based on the single-crystal data (CCDC-675375). The asymmetric unit of $Zn_2(bim)_4$ is also presented to illustrate the coordination environment of Zn atoms. (**D**) Transmission electron microscopy (TEM) image of $Zn_2(bim)_4$ MSNs. The inset shows the Tyndall effect of a colloidal suspension. (**E**) Illustration of the grid-like structure of the $Zn_2(bim)_4$ MSN. The Zn coordination polyhedra are depicted in blue, whereas the bim links are represented by sticks. H atoms are omitted for clarity. Symmetry code: A 2–*x*, *y*, 1.5–*z*. (**F**) Space-filling representation of a four-membered ring of the $Zn_2(bim)_4$ MSN.

of the single-layered nanosheet by HRTEM. The inset in Fig. 2A shows the cross-sectional HRTEM image of a five-layered $Zn_2(bim)_4$ nanosheet. The distance between the adjacent layers is ~1.18 nm, which is close to the d_{002} of the pristine crystals. The selected-area electron diffraction (SAED) pattern of a few-layered nanosheet collected along the *c* axis is shown in Fig. 2B, which corresponds to the diffraction pattern of the (*hk*0) planes of the nanosheet. The simulated SAED pattern is in good agreement with the experimental result, confirming that the exfoliated nanosheets are highly crystalline and retain the same in-plane crystal structure of pristine bulk materials. In Fig. 2C, a tapping-mode atomic force microscopy (AFM) image shows a $Zn_2(bim)_4$ MSN with large lateral dimension up to 1.5 μm. The height profile reveals that the nanosheet has a fairly flat, smooth terrace with a uniform thickness of 1.12 nm. Noting that 2D materials are often raised by a few angstroms above the supporting surface (*23*), the nanosheets in Fig. 2C should be monolayers of $Zn_2(bim)_4$. Additionally, it should be noted that $Zn_2(bim)_4$ MSNs with lateral size up to several microns were occasionally observed by TEM (fig. S4) and AFM (fig. S5). The exfoliation of layered $Zn_2(bim)_4$ precursors into nanosheets leads to a significant increase in the Brunauer-Emmett-Teller (BET) surface area, from 19.9 to 112.4 m^2/g (Fig. 2D). According to the classification by IUPAC (*24*), the isotherm of $Zn_2(bim)_4$ nanosheets can be identified as type II with an identifiable H4-type hysteresis loop, which is associated with the slit-like pores formed by aggregation of nanosheets during freeze drying. The Fourier transform infrared (FTIR) spectrum of $Zn_2(bim)_4$ nanosheets is identical to that of pristine $Zn_2(bim)_4$, indicating that the soft-physical exfoliation process has little effect on the in-plane structure of the $Zn_2(bim)_4$ nanosheets (fig. S6). Thermal analysis revealed that the $Zn_2(bim)_4$ nanosheets remain stable up to 200°C despite their nanometer-scale thickness (fig. S7).

Filtration of nanosheet suspensions through porous supports has been successfully applied to fabricate separating membranes in previous studies (*10*, *12*). This method, however, could not be applied in the current case. We attribute this to the fast restacking of nanosheets back to ordered pristine structures at elevated concentration when the solvent is filtered out (fig. S8). Ordered restacking of the MSNs will result in partial or total blockage of the molecular sieve pores. A hot-drop coating process addressed this problem, with the aim of achieving a disordered stacking of the nanosheets in the membrane layer. In a typical preparation, 3 ml of an MSN suspension was diluted five times with methanol and then deposited dropwise onto the surface of an α-Al_2O_3 disk (Inocermic GmbH, pore size 70 nm, Fig. 3A), which was heated at 120°C on a heating plate (*19*) (fig. S9). A uniform surface coverage of $Zn_2(bim)_4$ MSNs on α-Al_2O_3 support was achieved (fig. S10). Figure 3B presents a top-view SEM image of a $Zn_2(bim)_4$ MSN membrane at high magnification. Figure 3C shows the cross-sectional view of a membrane, where an unsupported portion of the nanosheet layer folds around the corner of the fractured support, clearly illustrating the high flexibility of $Zn_2(bim)_4$ nanosheets. In Fig. 3, B and C, the texture of the underlying α-Al_2O_3 support is distinguishable, indicating a very thin layer of $Zn_2(bim)_4$ MSNs on the support. In addition, a substantial aluminum signal from the underlying α-Al_2O_3 support was detected by x-ray photoelectron spectroscopy (XPS) for a $Zn_2(bim)_4$ MSN membrane, indicating that the MSN layer is only several nanometers thick (fig. S11).

The MSN membranes were sealed into Wicke-Kallenbach cells for measuring the permeance and selectivity of both single and binary feed of H_2 and CO_2. There was no absolute pressure differential across the membrane to avoid deforming the thin MSN layers (*19*) (fig. S12). We first conducted a control experiment using the α-Al_2O_3 support. The support showed an H_2 permeance of >50,000 GPUs with a low H_2/CO_2 selectivity (~1.2). Figure 3D shows the H_2/CO_2 selectivities measured on 15 membranes prepared via hot-drop coating at different support surface temperatures (80° to 200°C; table S1). In our investigation, 120°C optimizes the average selectivity and reproducibility. We hypothesize that this temperature is just high enough to maintain a solvent evaporation rate that avoids ordered restacking of the nanosheets during the coating process, which occurs at lower temperature (fig. S13). At higher temperature, pinholes or gaps might form between nanosheets when the solvent evaporates too quickly. The three membranes prepared at 120°C exhibited H_2/CO_2 selectivity of 261 ± 39. Nine membranes were prepared at 120°C with different drop volumes (1 to 15 ml), which proved to be a less significant factor than the support surface temperature (fig. S14 and table S2).

A membrane prepared at 120°C was tested for single gas permeation of H_2 (0.29 nm), CO_2 (0.33 nm), N_2 (0.36 nm), CH_4 (0.38 nm), and C_2H_6 (0.44 nm). There was a clear cutoff between H_2 and CO_2 (fig. S15). H_2/CO_2 separation measurements at different temperatures indicate activated diffusion for both H_2 and CO_2 (fig. S16). These results suggest a size exclusion mechanism for separation of H_2 from CO_2. H_2/CO_2 separation with different feed compositions indicates that

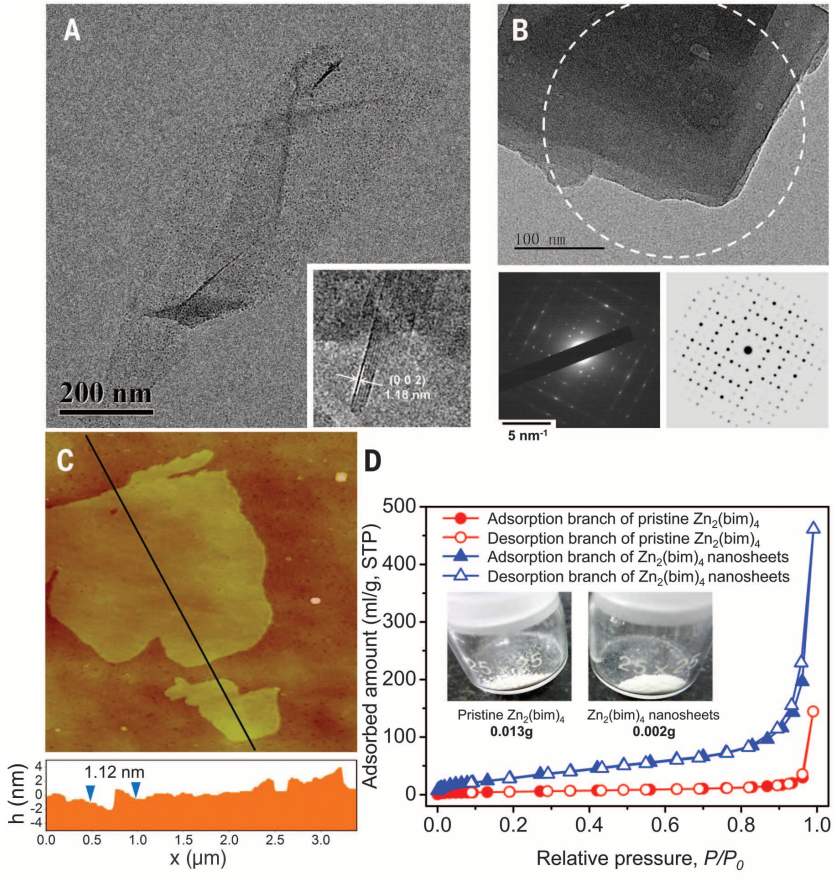


Fig. 2. Characterizations of molecular sieve nanosheets. (**A**) Low-magnification TEM image of a piece of $Zn_2(bim)_4$ nanosheets. The inset image shows a high-resolution TEM image of a five-layered $Zn_2(bim)_4$ nanosheet. (**B**) SAED pattern (white circle) shows the diffraction from (hk0) planes within a few-layered nanosheet. A simulated SAED pattern of $Zn_2(bim)_4$ nanosheet down the *c* axis is also shown. (**C**) Tapping-mode AFM topographical image of $Zn_2(bim)_4$ nanosheets on silicon wafer. The height profile of the nanosheets along the black lines was marked in the image. (**D**) N_2 adsorption-desorption isotherms (77 K) on pristine $Zn_2(bim)_4$ and $Zn_2(bim)_4$ nanosheets. The inset presents photographs comparing the pristine $Zn_2(bim)_4$ and $Zn_2(bim)_4$ nanosheets obtained after exfoliation.

CO_2 adsorption has influence on the H_2 permeation to some extent (fig. S17).

Normally, researchers prepare thicker membranes to minimize nonselective defects. However, there is an inverse relationship between membrane permeance and selectivity. In contrast, we found an anomalous proportional relationship between permeance and selectivity for the $Zn_2(bim)_4$ MSN membranes, as shown in Fig. 3E. The H_2 permeance varies from 760 to 3760 GPUs along with the variation of H_2/CO_2 selectivity from 53 to 291. The linear regression of these data accounts for 90.6% of the variance (table S1). Meanwhile, a small fluctuation in CO_2 permeance (12.3 ± 4.5 GPUs) was measured on all 15 membranes (table S1). Theoretically, only H_2 molecules can pass through the four-membered rings of the MSNs. The non-zero CO_2 permeance could be attributed to the non–size-selective mass transport through imperfect sealing or through boundaries of the nanosheet layer.

To gain insight into the structure-performance relationship of the MSN membranes, four membranes with different performance were characterized by powder XRD (Fig. 3F). The proportional relationship between permeance and selectivity can be observed in the table in Fig. 3F. Membranes with higher permeance are more selective, and vice versa. Membrane performance can be correlated with the order of nanosheet stacking, as indicated by XRD patterns. The low-angle hump suggests regions with expanded stacking of nanosheets. The appearance of (002) reflection coincides with the formation of bulk $Zn_2(bim)_4$ via ordered restacking. Lamellar ordering of nanosheets would block the permeation pathway for H_2, but have only a slight effect on CO_2 leakage, resulting in a reduction in membrane performance. Noting that a tiny hump is still identifiable for the high-performance membrane, we speculate that an ultrapermeable and superselective molecular sieve membrane could be obtained if the membrane consisted of fully disordered MSNs.

The performance of our membranes exceeds the latest Robeson's upper-bound for H_2/CO_2 gas pair and is higher than that of the molecular sieve membranes reported to date, including polycrystalline membranes composed of three dimensional MOFs (fig. S18). For practical use, thermal and hydrothermal stability is an important issue for H_2 selective membranes (*25*). A $Zn_2(bim)_4$ MSN membrane was tested under different conditions for more than 400 hours in total, including two temperature cycles, showing no degradation in membrane performances. When exposed to an equimolar H_2/CO_2 feed containing ~4 mol % steam at 150°C, the membrane showed good stability after a 120-hour test (fig. S19).

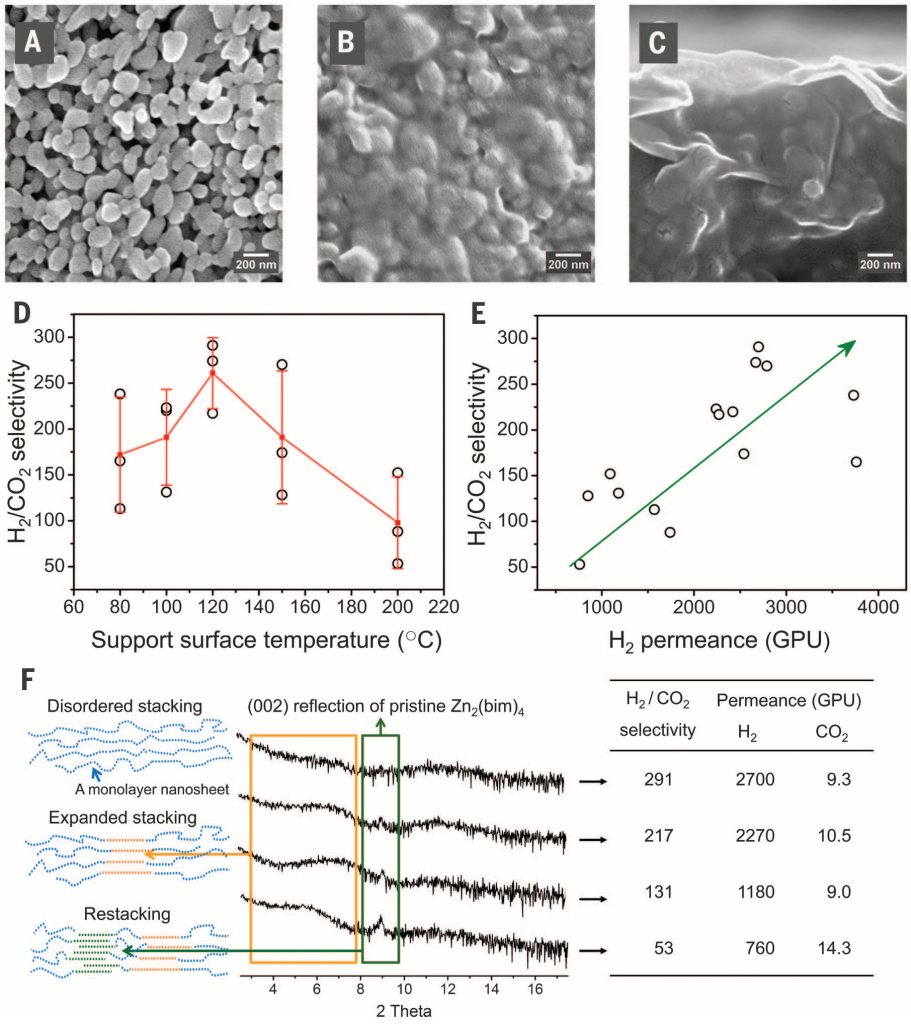


Fig. 3. Morphology and performance of membranes derived from molecular sieve nanosheets. (**A**) SEM image of a bare porous α-Al_2O_3 support. (**B**) SEM top view and (**C**) cross-sectional view of a $Zn_2(bim)_4$ nanosheet layer on α-Al_2O_3 support. (**D**) Scatterplot of H_2/CO_2 selectivities measured from 15 membranes. The red line with symbols shows the average selectivity and dispersion of selectivity of the membranes prepared at different coating temperatures. (**E**) Anomalous relationship between selectivity and permeance measured from 15 membranes. (**F**) Powder XRD patterns of four membranes with different separation properties. The cartoons at the left schematically illustrate the microstructural features of the nanosheet layers. The yellow and green portions correspond to the low-angle humps and the (002) peaks in the XRD patterns, respectively. All the membranes were measured for equimolar mixtures at room temperature and 1 atm.

REFERENCES AND NOTES

1. L. M. Robeson, *J. Membr. Sci.* **320**, 390–400 (2008).
2. D. L. Gin, R. D. Noble, *Science* **332**, 674–676 (2011).
3. R. M. de Vos, H. Verweij, *Science* **279**, 1710–1711 (1998).
4. M. B. Shiflett, H. C. Foley, *Science* **285**, 1902–1905 (1999).
5. Z. Lai *et al.*, *Science* **300**, 456–460 (2003).
6. H. B. Park *et al.*, *Science* **318**, 254–258 (2007).
7. M. Carta *et al.*, *Science* **339**, 303–307 (2013).
8. R. K. Joshi *et al.*, *Science* **343**, 752–754 (2014).
9. W. Kim, S. Nair, *Chem. Eng. Sci.* **104**, 908–924 (2013).
10. K. Varoon *et al.*, *Science* **334**, 72–75 (2011).
11. W. J. Roth, P. Nachtigall, R. E. Morris, J. Čejka, *Chem. Rev.* **114**, 4807–4837 (2014).
12. H. Li *et al.*, *Science* **342**, 95–98 (2013).
13. Y. S. Lin, I. Kumakiri, B. N. Nair, H. Alsyouri, *Sep. Purif. Rev.* **31**, 229–379 (2002).
14. A. K. Cheetham, C. N. R. Rao, R. K. Feller, *Chem. Commun.* **46**, 4780–4795 (2006).
15. P. Z. Li, Y. Maeda, Q. Xu, *Chem. Commun.* **47**, 8436–8438 (2011).
16. A. Gallego *et al.*, *Adv. Mater.* **25**, 2141–2146 (2013).
17. J. C. Tan, P. J. Saines, E. G. Bithell, A. K. Cheetham, *ACS Nano* **6**, 615–621 (2012).
18. N. W. Ockwig, T. M. Nenoff, *Chem. Rev.* **107**, 4078–4110 (2007).
19. See supplementary materials on *Science* Online.
20. Q. F. Yang *et al.*, *CrystEngComm* **10**, 1534–1541 (2008).
21. K. S. Park *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 10186–10191 (2006).
22. S. C. Junggeburth *et al.*, *J. Am. Chem. Soc.* **135**, 6157–6164 (2013).
23. K. S. Novoselov *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 10451–10453 (2005).
24. K. S. W. Sing Ed, *Pure Appl. Chem.* **54**, 2201–2218 (1982).
25. H. B. Wang, X. L. Dong, Y. S. Lin, *J. Membr. Sci.* **450**, 425–432 (2014).

ACKNOWLEDGMENTS

We thank the National Natural Science Foundation of China and the Key Research Program of the Chinese Academy of Sciences for funding, L. Liu for experimental assistance in wet ball milling, S. Miao for experimental assistance in TEM and SAED, and H. Duan for experimental assistance in TG analysis. Y.P., Y.L., and W.Y. are inventors on a Chinese patent filed through the Dalian Institute of Chemical Physics (CN102974229A, "Exfoliation and application of two-dimensional layered metal-organic frameworks").

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/346/6215/1356/suppl/DC1
Materials and Methods
Figs. S1 to S19
Tables S1 and S2
References (*26*–*40*)

1 April 2014; accepted 12 November 2014
10.1126/science.1254227

POLLINATOR DECLINES

Extinctions of aculeate pollinators in Britain and the role of large-scale agricultural changes

Jeff Ollerton,[1]* Hilary Erenler,[1] Mike Edwards,[2] Robin Crockett[1]

Pollinators are fundamental to maintaining both biodiversity and agricultural productivity, but habitat destruction, loss of flower resources, and increased use of pesticides are causing declines in their abundance and diversity. Using historical records, we assessed the rate of extinction of bee and flower-visiting wasp species in Britain from the mid-19th century to the present. The most rapid phase of extinction appears to be related to changes in agricultural policy and practice beginning in the 1920s, before the agricultural intensification prompted by the Second World War, often cited as the most important driver of biodiversity loss in Britain. Slowing of the extinction rate from the 1960s onward may be due to prior loss of the most sensitive species and/or effective conservation programs.

Pollinating insects, particularly bees and other flower-visiting Hymenoptera (Aculeata), are some of the most ecologically and economically important insects (*1*–*3*) but have declined in species richness, geographical range, and abundance (*2*–*5*). Previous studies have assessed the roles played by habitat destruction and loss of flower resources (*4*, *5*), as well as pesticides (*6*), over relatively modest time scales and geographical ranges. Analyses of regions are rare (*7*–*10*), and our understanding of the effects of human-mediated actions over longer periods is limited. Here we assess the bee and flower-visiting wasp species that have gone extinct in Britain, using 494,117 records held by the Bees, Wasps and Ants Recording Society (BWARS), probably the most detailed available for a single country. We define extinct species as those that have not been recorded for at least 20 years after their last observation, despite extensive efforts by members of BWARS and other naturalists.

Twenty-three bee and flower-visiting wasp species have become extinct in Britain (Table 1), including formerly widespread species. We exclude single early records that cannot be verified as representing stable breeding populations, but we include one species that has recolonized Britain after an absence of six decades (see materials and methods and other supplementary materials).

Since the mid-19th century, the pattern of British bee and wasp extinctions has been characterized by intervals of relative stability in which few species were lost, interspersed with times when more than three species per decade went extinct (Fig. 1 and Table 2). These data indicate a period of relatively sustained extinctions from the late 1920s to the late 1950s, with other isolated extinction peaks before and after this time. These features are confirmed in Fig. 2, where the average gradient indicates the relative extinction rate over a period, and the period of sustained extinctions is evident as the phase of maximum gradient during the mid-20th century.

The varying rates of extinctions were quantified by applying breakpoint analysis to the cumulative record. In this analysis, a piecewise linear model is fitted to data to reveal periods of approximately constant extinction rate, separated by breakpoints where the rate changes. The analysis was iterated for up to 10 breakpoints, and the Akaike information criterion (AIC), confirmed by coefficient of determination (multiple R^2), was used to establish the best model (see supplementary materials). For these data, changes in AIC and multiple R^2 level off for two models having four breakpoints (table S2). These are very similar, sharing the latter three breakpoints and revealing effectively identical periods of approximately uniform extinction rate for the majority of the 20th century (Table 2).

Both models must be interpreted with caution, as the data for "year last recorded" may not equate to "year last living." Declines in populations due to habitat changes may indicate that a species went unrecorded for some years before the actual extinction. The robustness of the breakpoints to this potential ambiguity of the probability of the year last living has been assessed, and though there is some sensitivity in the timing of the earlier and later breakpoints, due to the sparseness and bunching of events at the ends of the record, the period of sustained extinctions from the late 1920s to the late 1950s is very stable. We also assessed how variability in recorder effort over time may have affected our findings, using the number of records per decade in the BWARS database as a proxy for effort, and found that our results were not systematically affected by this. These analyses are discussed in the supplementary materials.

Table 1. Extinct British bee and flower-visiting wasp species, ordered by their last observed year, with number of records of that species from the BWARS database. A record is defined as an occurrence of a species on a specific date, at a location, and by a specific person. Some of the earlier records relate to larger geographic areas over longer time periods (e.g., presence of a species in a county in a year), whereas later records are at particular grid references.

Species, naming authority, and date described	Number of BWARS records	Year last observed
Lestica clypeata (Schreber 1759)	2	1853
Psen ater (Olivier 1792)	2	Pre-1880
Dufourea minuta Lepeletier 1841	3	1881
Odynerus reniformis (Gmelin 1790)	17	1909
Philocetes truncatus (Dahlbom 1831)	2	1910
Melecta luctuosa (Scopoli 1770)	16	1912
Halictus maculatus Smith 1848	26	1930
Andrena nana (Kirby 1802)	4	1930
Andrena polita Smith 1847	11	1934
Arachnospila rufa (Haupt 1927)	2	1938
Bombus cullumanus (Kirby 1802)	22	1941
Andrena tridentata (Kirby 1802)	6	1944
Andrena vaga Panzer 1799	3	1946
Mellinus crabroneus (Thunberg 1791)	26	1952
Andrena lepida Schenk 1861	2	1952
Dufourea halictula (Nylander 1852)	6	1953
Chrysis longula Abeille de Perrin 1879	6	1957
Ancistrocerus quadratus (Panzer 1799)	8	1968
Eucera nigrescens Perez 1879	26	1970
Bombus subterraneus Linnaeus 1758	268	1988
Ancistrocerus antilope (Panzer 1798)	24	1989
Chrysis pseudobrevitarsis Linsenmaier 1951	3	1989
Andrena lathyri Alfken 1899	11	1990

[1]Department of Environmental and Geographical Sciences, School of Science and Technology, University of Northampton, Avenue Campus, Northampton NN2 6JD, UK. [2]Lea-Side, Carron Lane, Midhurst GU29 9LB, UK.
*Corresponding author. E-mail: jeff.ollerton@northampton.ac.uk

Downloaded from www.sciencemag.org on December 14, 2014

Some of the phases of acceleration in the rate of species loss coincide with large-scale changes in agricultural policy and practice in Britain. For example, the second half of the 19th century saw the increased import of South American guano as soil fertilizer (*11*), which had a double impact on bee and wasp floral resources: (i) increased grass productivity at the expense of wild flower diversity (*12*); and (ii) decline in reliance on strict rotational cropping. The latter would have included fallow years, in which nectar-rich weeds flourished, and a legume rotation offering resources favored by long-tongued bees (*13*). Additionally, during the late 19th and early 20th centuries, the area of arable and fodder crops declined by more than 55%, replaced by permanent grassland (*11*). After World War I, food security concerns led to agricultural reforms that further intensified farming in Britain. This was aided by the invention of the Haber process, allowing the industrial manufacture of inorganic nitrogen fertilizers (*14*) and probably accelerating the decline of wild flowers (*12*). This succession of events correlates in time with the first two phases in the extinction record, as shown by both models, up to the late 1920s (Table 2), characterized by extinction rates rising from 0.21 species per decade in the 1850s–1870s to 1.31 in the 1900s–1920s.

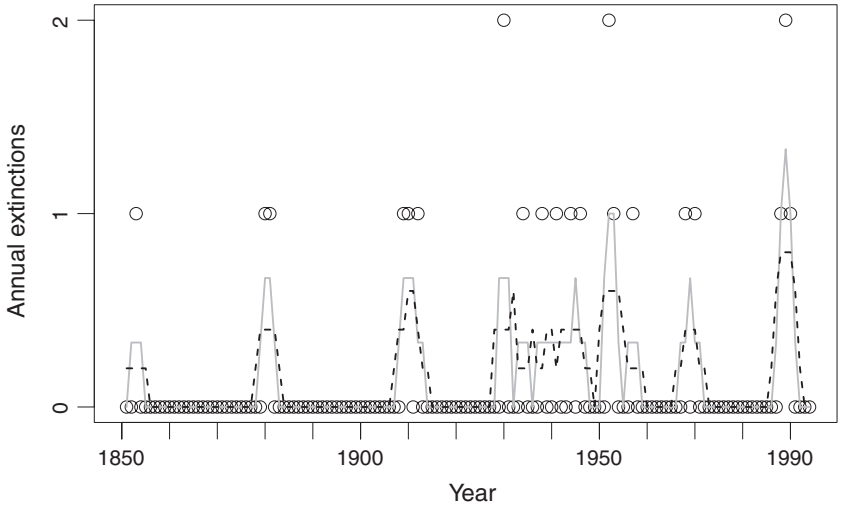


Fig. 1. Annual British bee and flower-visiting wasp extinctions, 1851–1994. The numbers of annual species extinctions are shown as black circles; 3- and 5-year moving-averaged annual extinction data are shown as solid gray and dashed black lines, respectively.

Table 2. Decadal extinction rates of British bees and flower-visiting wasps during the five time periods defined by the four breakpoints. Set-type 1 is the most probable, as represented in Fig. 2; set-type 2 is the four-breakpoint set with a slightly better fit, as defined by AIC.

Period	Decadal extinction rate			Multiple R^2
	Estimate	Lower 99% CI	Upper 99% CI	
		Set-type 1		
Up to 1874	0.21	−0.02	0.44	0.239
1874–1928	0.96	0.80	1.12	0.832
1928–1958	3.46	3.14	3.78	0.969
1958–1986	0.98	0.66	1.30	0.740
1986 onward	5.48	0.05	10.91	0.700
		Set-type 2		
Up to 1902	0.61	0.49	0.73	0.784
1902–1929	1.31	0.79	1.82	0.665
1929–1959	3.41	3.07	3.74	0.967
1959–1986	0.98	0.66	1.30	0.740
1986 onward	5.48	0.05	10.91	0.700

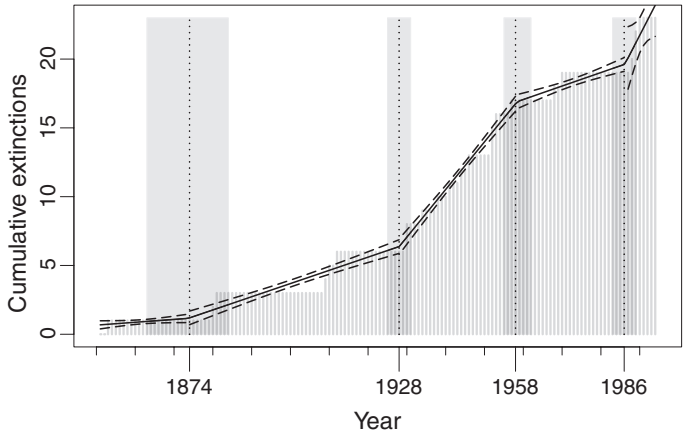


Fig. 2. Cumulative British bee and flower-visiting wasp extinctions, 1851–1994. Data are plotted as cumulative number of extinctions per year (vertical gray bars). The four estimated breakpoints are shown as vertical dotted black lines, with 99% CIs depicted as transparent gray-shaded rectangles. The identified line segments are given by solid black lines, with 99% CIs indicated by dashed lines. The smaller horizontal axis tick marks show decades starting at 1850.

The third phase from the late 1920s to the late 1950s can be attributed to agricultural intensification after World War I and during and after World War II, marking the greatest loss of bees and wasps at 3.41 to 3.46 species per decade (Table 2). The period from the late 1950s to the mid-1980s showed a slowing of the extinction rate to ~0.98 species per decade (Table 2), which is not easily explained in light of intensification of farming encouraged by Common Agricultural Policy subsidies. Improvement of land previously deemed uneconomic for production resulted in further losses of pollinator habitats such as hedgerows and species-rich grassland (*15*), so slowing may be a result of the most sensitive species having been already lost, or it may have occurred because conservation initiatives are working.

The final period from 1986 to 1994 could be seen as contradicting recent evidence of a slowing rate of pollinator decline in northwest Europe (*9*), but this should be interpreted cautiously. The high calculated extinction rate and its large confidence interval (CI) (Fig. 2) arise because of the four 1988–1990 extinctions in the otherwise zero-extinction period from 1971–1994. In addition, the provisional 1995–2013 record reveals no extinctions. If the passage of time confirms this record, then the four extinctions will form an isolated cluster in a zero-extinction period extending from 1971; otherwise, they could mark the start of a further period of high extinction rate (see supplementary materials).

Our study adds to a debate on the rates and causes of regional and country-wide extinctions of British biodiversity (including invertebrates, vertebrates, and plants) and the limitations imposed by data quality [e.g., (*16*–*20*)]. The available data for bee and flower-visiting wasp extinctions within Britain show that there are deep historical roots to this loss in pollinator diversity that correlate with transformations of land management related to changes in agricultural policy and practice, a conclusion also drawn by these

other studies (*16*–*20*). Agriculture accounts for 70% of British land use, strongly suggesting that this relationship is causal, though the exact drivers of extinctions are clearly multifactorial and complex. For example, for some species there may have been a mismatch in the timing of extinctions in relation to specific agricultural changes (an "extinction debt") that we cannot currently identify.

Finally, we note that the United Kingdom is on the northern and western edge of the distribution range for many Hymenoptera, resulting in the recent colonization of species that had not previously been recorded, such as *Bombus hypnorum* (*21*) and *Colletes hederae* (*22*). We might therefore expect other colonizations, extirpations, and recolonizations as part of normal background ecological processes, regardless of human activity (see supplementary materials). The consequences of climate change on species distributions provides further complications, and disentangling anthropogenic versus natural effects poses a future challenge for researchers.

REFERENCES AND NOTES

1. J. Ollerton, R. Winfree, S. Tarrant, *Oikos* **120**, 321–326 (2011).
2. S. G. Potts *et al.*, *Trends Ecol. Evol.* **25**, 345–353 (2010).
3. T. D. Breeze, A. P. Bailey, K. G. Balcombe, S. G. Potts, *Agric. Ecosyst. Environ.* **142**, 137–143 (2011).
4. J. C. Biesmeijer *et al.*, *Science* **313**, 351–354 (2006).
5. C. Carvell *et al.*, *Biol. Conserv.* **132**, 481–489 (2006).
6. H. C. Godfray *et al.*, *Proc. R. Soc. London Ser. B* **281**, 20140558 (2014).
7. L. A. Burkle, J. C. Marlin, T. M. Knight, *Science* **339**, 1611–1615 (2013).
8. I. Bartomeus *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 4656–4660 (2013).
9. L. G. Carvalheiro *et al.*, *Ecol. Lett.* **16**, 870–878 (2013).
10. C. L. Morales, M. P. Arbetman, S. A. Cameron, M. A. Aizen, *Front. Ecol. Environ* **11**, 529–534 (2013).
11. P. Brassley, *Agric. Hist. Rev.* **48**, 60–84 (2000).
12. D. A. Wedin, D. Tilman, *Science* **274**, 1720–1723 (1996).
13. D. Vanengelsdorp, M. D. Meixner, *J. Invertebr. Pathol.* **103** (suppl. 1), S80–S95 (2010).
14. V. Smil, *Enriching the Earth: Fritz Haber, Carl Bosch, and the Transformation of World Food Production* (The MIT Press, Cambridge, MA, 2001).
15. R. A. Robinson, W. J. Sutherland, *J. Appl. Ecol.* **39**, 157–176 (2002).
16. C. Hambler, M. R. Speight, *Conserv. Biol.* **10**, 892–896 (1996).
17. C. Hambler, P. A. Henderson, M. R. Speight, *Biol. Conserv.* **144**, 713–721 (2011).
18. D. McCollin, L. Moore, T. Sparks, *Biol. Conserv.* **92**, 249–263 (2000).
19. J. A. Thomas *et al.*, *Science* **303**, 1879–1881 (2004).
20. C. Hambler, M. R. Speight, *Science* **305**, 1563–1565, author reply 1563–1565 (2004).
21. D. Goulson, P. H. Williams, *British J. Ent. Nat. Hist.* **14**, 129–131 (2001).
22. Bees, Wasps and Ants Recording Society species accounts: *Colletes hederae* Schmidt and Westrich, 1993; www.bwars.com/index.php?q=bee/colletidae/colletes-hederae (2011).

ACKNOWLEDGMENTS

We thank W. E. Kunin and two referees for valuable comments and BWARS for permission to use its data, which are available on the UK's National Biodiversity Network Gateway: https://data.nbn.org.uk/Datasets/GA000055.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/346/6215/1360/suppl/DC1
Materials and Methods
Tables S1 to S4
References (*23*–*25*)

10 June 2014; accepted 5 November 2014
10.1126/science.1257259

NUTRITIONAL IMMUNITY

Escape from bacterial iron piracy through rapid evolution of transferrin

Matthew F. Barber and Nels C. Elde*

Iron sequestration provides an innate defense, termed nutritional immunity, leading pathogens to scavenge iron from hosts. Although the molecular basis of this battle for iron is established, its potential as a force for evolution at host-pathogen interfaces is unknown. We show that the iron transport protein transferrin is engaged in ancient and ongoing evolutionary conflicts with TbpA, a transferrin surface receptor from bacteria. Single substitutions in transferrin at rapidly evolving sites reverse TbpA binding, providing a mechanism to counteract bacterial iron piracy among great apes. Furthermore, the C2 transferrin polymorphism in humans evades TbpA variants from *Haemophilus influenzae*, revealing a functional basis for standing genetic variation. These findings identify a central role for nutritional immunity in the persistent evolutionary conflicts between primates and bacterial pathogens.

Iron is a precious cellular metal, sequestered by hosts and scavenged by pathogens (*1*–*3*). Vertebrate iron transport is mediated by serum transferrin, a protein that binds circulating iron and delivers it to cells via receptor-mediated endocytosis. Modern transferrin arose through a tandem duplication event in ancestral metazoans that produced two homologous domains, the N and C lobes, each of which binds a single iron ion with high affinity (*4*). Transferrin also contributes to host nutritional immunity by sequestering essential iron away from microbial pathogens. One hallmark of host immunity protein evolution is recurrent positive selection driven by diverse and rapidly evolving viruses (*5*). However, the essential nature of transferrin's role in iron transport necessitates functional conservation, which may impede adaptation against iron piracy. Indeed, the effect of nutritional immunity on evolution at host-pathogen interfaces is unclear.

To determine whether transferrin might be subject to pathogen-driven evolution in the primate lineage, we cloned and sequenced transferrin orthologs from 21 hominoid, Old World, and New World monkey species for phylogenetic analysis (Fig. 1A and fig. S1). A combination of maximum likelihood–based algorithms to assess ratios of nonsynonymous to synonymous substitution rates (dN/dS) revealed strong signatures of episodic positive selection in transferrin ($P < 0.0001$) (table S1) across several branches of the primate lineage (Fig. 1A, fig. S2, and tables S2 to S8). To date, such signatures of molecular "arms races" in mammals are primarily documented among cell surface receptors and innate pattern recognition proteins antagonized by viruses (*6*–*8*), reflecting the primacy of such host proteins as "front line" immune defenses or points of entry for viruses. These results indicate that primate transferrin has undergone bouts of rapid evolution reminiscent of canonical innate immunity factors engaged in host-pathogen arms races.

Our analysis of positive selection in transferrin revealed that 16 of 18 rapidly evolving sites map to the C lobe (Fig. 1B), despite the fact that the N and C lobes are functionally homologous for iron binding and transport. This contrast was particularly clear in the hominoid lineage, where the C lobe alone shows strong evidence of positive selection ($P < 0.0001$), whereas the N lobe does not ($P > 0.99$) (table S4) and instead has evolved under purifying selection. The transferrin N and C lobes have thus been subject to very different selective pressures during their respective evolutionary histories, despite performing identical essential physiologic functions. Previous reports indicate that the transferrin receptor (Tf-R) in rodents and carnivores has been subject to positive selection driven by viral entry proteins (*6*, *9*). However, Tf-R is subject to purifying selection in primates (fig. S3A), and only one of 18 rapidly evolving sites in transferrin makes contact with Tf-R (fig. S3B), indicating that, as expected, rapid evolution of transferrin has not been driven by coevolution with its cognate receptor.

Remarkably, 14 of 16 rapidly evolving sites in the transferrin C lobe form direct contacts with transferrin binding protein A (TbpA) from *Neisseria meningitidis* when mapped to a recently solved, high-resolution co-crystal structure of human transferrin bound to TbpA (Fig. 1, B and C, and fig. S4) (*10*). Several Gram-negative human pathogens, including *Neisseria gonorrhoeae*, *N. meningitidis*, and *Haemophilus influenzae*, scavenge host iron via surface receptors that bind and extract iron exclusively from the C lobe of transferrin (*11*–*13*). As the primary component of these bacterial receptors, TbpA is a transmembrane transporter that facilitates extraction and translocation of iron into the bacterial periplasm. Notably, the specificity of TbpA proteins for their respective host transferrin is hypothesized to restrict the host range of these bacteria (*14*, *15*). The rapidly evolving sites that

Department of Human Genetics, University of Utah School of Medicine, Salt Lake City, UT 84112, USA.
*Corresponding author. E-mail: nelde@genetics.utah.edu

Downloaded from www.sciencemag.org on December 14, 2014

we identified cluster within six of the seven subregions of the TbpA interface, proximal to the TbpA L3 helix finger and plug loop that have been defined as critical points of contact between these two proteins (*10*). In contrast, no rapidly evolving sites in transferrin form contacts with TbpB, a bacterial accessory co-receptor that coordinately binds the C lobe of transferrin for iron acquisition (*10*, *16*) (fig. S5). Thus, positive selection of the transferrin C lobe may have been solely driven by interactions with TbpA and ancestral TbpA-like proteins.

To define the functional consequences of rapid evolution in primate transferrin, we used competitive binding assays (see the supplementary materials) to directly assess interactions between transferrin and TbpA (*17*, *18*). For these experiments, we purified recombinant transferrin proteins from a panel of primate species (fig. S6) and expressed TbpA from the human-specific pathogens *N. gonorrhoeae* and *H. influenzae* (strains MS11 and Eagan, respectively) in nonpathogenic BL21 *Escherichia coli*, which do not possess transferrin receptors (fig. S7). Consistent with previous studies, TbpA from both pathogens bound strongly to recombinant human transferrin (*17*, *19*) (Fig. 1D). TbpA variants also recognized transferrin from gorillas but not chimpanzees, orangutans, gibbons, or baboons (Fig. 1D). The lack of binding between TbpA and chimpanzee transferrin was particularly striking given that chimpanzees represent the closest primate relative to gorillas and humans. These observations were independently corroborated by competitive binding enzyme-linked immunosorbent assays (ELISA) recapitulating transferrin-TbpA interactions (Fig. 1D, bar graphs). Thus, transferrin divergence among great apes is sufficient to dictate distinct outcomes in bacterial TbpA binding interactions.

Among the four amino acid differences between human and chimpanzee transferrin, position 591

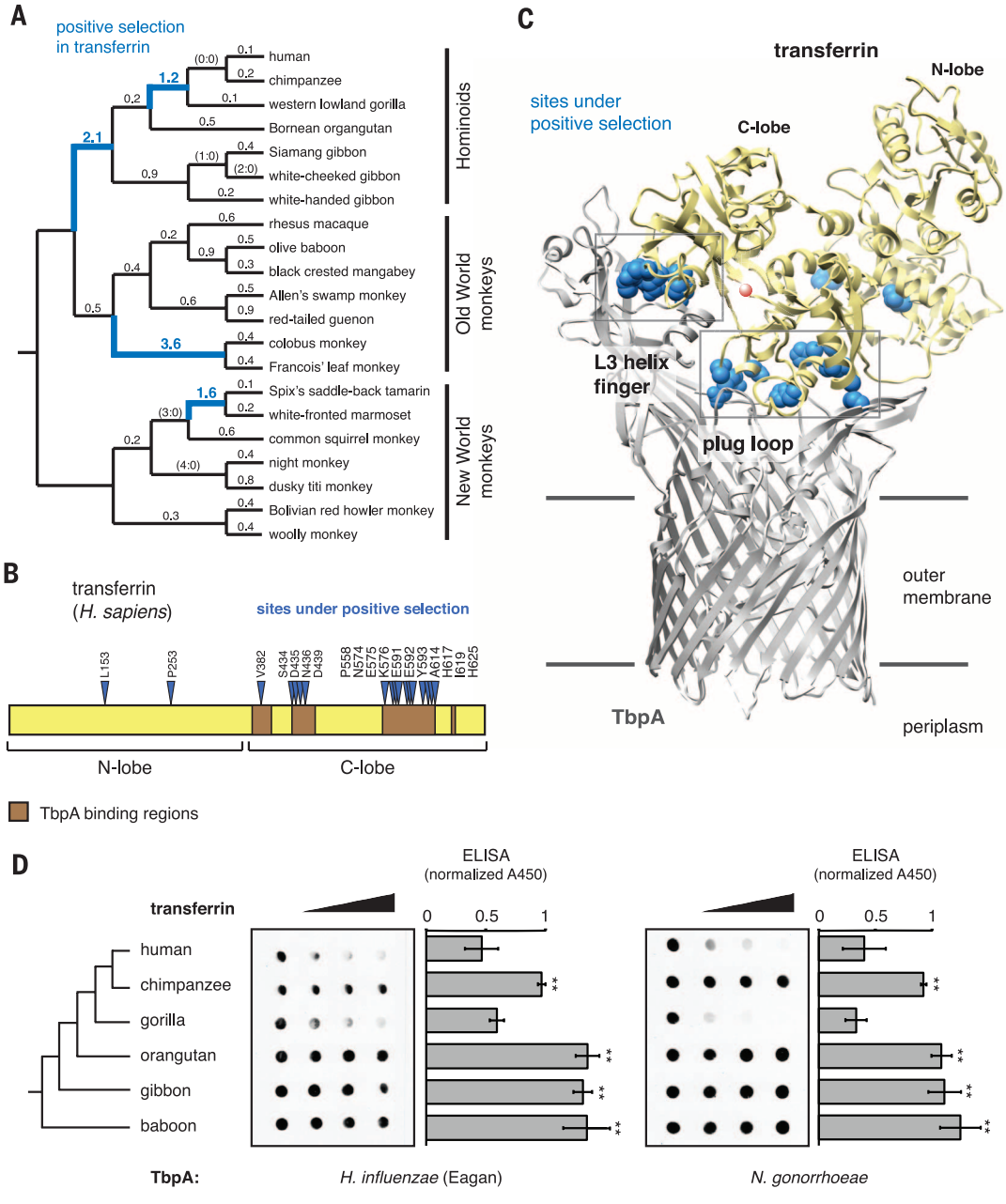


Fig. 1. Primate transferrin has undergone recurrent positive selection at the binding interface with bacterial TbpA. (**A**) A primate phylogram highlighting rapid evolution of transferrin. d*N*/d*S* ratios along each branch of the primate lineage are listed, with d*N*/d*S* values >1 highlighted in blue. Branches with no synonymous changes display N:S substitution ratios in parentheses. (**B**) Schematic of the human transferrin protein (yellow). Amino acid positions showing evidence of positive selection (phylogenetic analysis by maximum likelihood (*30*), posterior probability > 0.99) are denoted by blue arrows. Regions of transferrin spanning the *N. meningitidis* TbpA binding interface (*10*) are highlighted in brown. (**C**) Ribbons representation of the structure (Protein Data Bank no. 3V8X) of human transferrin (yellow) in complex with *N. meningitidis* TbpA (gray), with the position of the bacterial outer membrane labeled. Amino acid side chains in transferrin with strong evidence of positive selection are marked by blue spheres. The position of a ferric ion is modeled as a red dot. (**D**) Competitive TbpA binding assays performed as dot blot (left) or ELISA (right), using *E. coli* expressing indicated pathogen TbpA. Samples were incubated with horseradish peroxidase (HRP)–conjugated human transferrin alone (0.5 μg/mL), or HRP-transferrin in the presence of increasing concentrations of recombinant purified transferrin (5, 10, or 20 μg/mL) from indicated primates. Error bars represent SD of four independent experiments. ***P* < 0.01 relative to human transferrin.

(glutamic acid in humans, lysine in chimpanzees) shows strong signals of recurrent positive selection (Fig. 2A and table S6) and, intriguingly, also lies proximal to the interface with the L3 helix finger of TbpA, which plays a pivotal role in iron acquisition (*10*) (Fig. 2B and fig. S4). Competitive binding assays revealed that a glutamic acid to lysine substitution at this position in human transferrin (E591K) is sufficient to impair binding to TbpA from both *N. gonorrhoeae* and *H. influenzae*, rendering its binding affinity similar to chimpanzee transferrin (Fig. 2C). In addition, introducing a glutamic acid to lysine substitution at position 591 (K591E) in chimpanzee transferrin is sufficient to restore TbpA binding (Fig. 2C). Variation at position 591 thus directly links positive selection of transferrin to recognition by bacterial TbpA. Furthermore, closely related bonobos share glutamic acid 591 with humans and gorillas (Fig. 2A), providing clear evidence of recent transferrin adaptation in chimpanzees. These findings are surprising, given that previous studies concluded that the transferrin-TbpA interaction is largely impervious to single-point mutations in TbpA, likely owing to a substantial 2500 $Å^2$ binding interface (*10*). This result highlights the predictive power of positive selection analyses to pinpoint residues that most drastically alter binding affinity at host-pathogen interfaces. Our findings also bolster the hypothesis that transferrin is one of a handful of factors limiting the host range of human-specific bacterial pathogens (*15*, *20*), even among closely related primate species.

Given the ability of a single substitution to dictate TbpA recognition between primate transferrin orthologs, we were curious whether standing genetic variation of transferrin in human populations might provide similar protection. After the major C1 allele, C2 is the most abundant transferrin variant, found at roughly 6 to 26% allele frequency across human populations (*21*) (Fig. 2D and tables S9 and S10). C2 differs from C1 by a single C/T substitution that changes proline 589 to serine (Fig. 2D and table S9). Although this variation in transferrin has long been recognized, no appreciable differences in iron binding or other activities have been discerned between C1 and C2 (*22*). However, this polymorphism occurs only two amino acids away from position 591, which is sufficient to control TbpA binding (Fig. 2B, E). In competitive binding assays, the C2 variant was markedly resistant to recognition by TbpA from *H. influenzae* (Fig. 2F), providing a striking example of functionally adaptive consequences for genetic variation in humans. This result highlights the effect of nutritional immunity on primate evolution, from 40 million years of species divergence to a single polymorphism circulating in human populations.

To quantify differences in TbpA binding with transferrin human variants, we generated dissociation curves and calculated half-maximal inhibitory concentrations (IC_{50}). IC_{50} calculations revealed severely reduced binding by *H. influenzae* TbpA to C2 transferrin relative to C1 (fig. S8 and table S11). Unlike TbpA from *H. influenzae*, the *N. gonorrhoeae* variant bound C2 with nearly equal affinity to C1 transferrin as determined by competitive binding assays and IC_{50} calculations (Fig. 2F, fig. S8, and table S11), indicating functional variability among pathogen TbpA orthologs.

To delineate functional outcomes among TbpA proteins, we sampled variants from additional pathogenic strains isolated in clinics. We found that TbpA from *H. influenzae* isolates Eagan and strain 11 are specific for recognition of the C1 variant, potentially at the expense of C2 recognition (Fig. 2F and Fig. 3, A and B). In contrast, TbpA from *N. meningitidis* and *N. gonorrhoeae* displayed similar binding to both C1 and C2 transferrin, whereas *H. influenzae* strain 15 TbpA exhibited intermediate C2 binding affinity (Fig. 3,

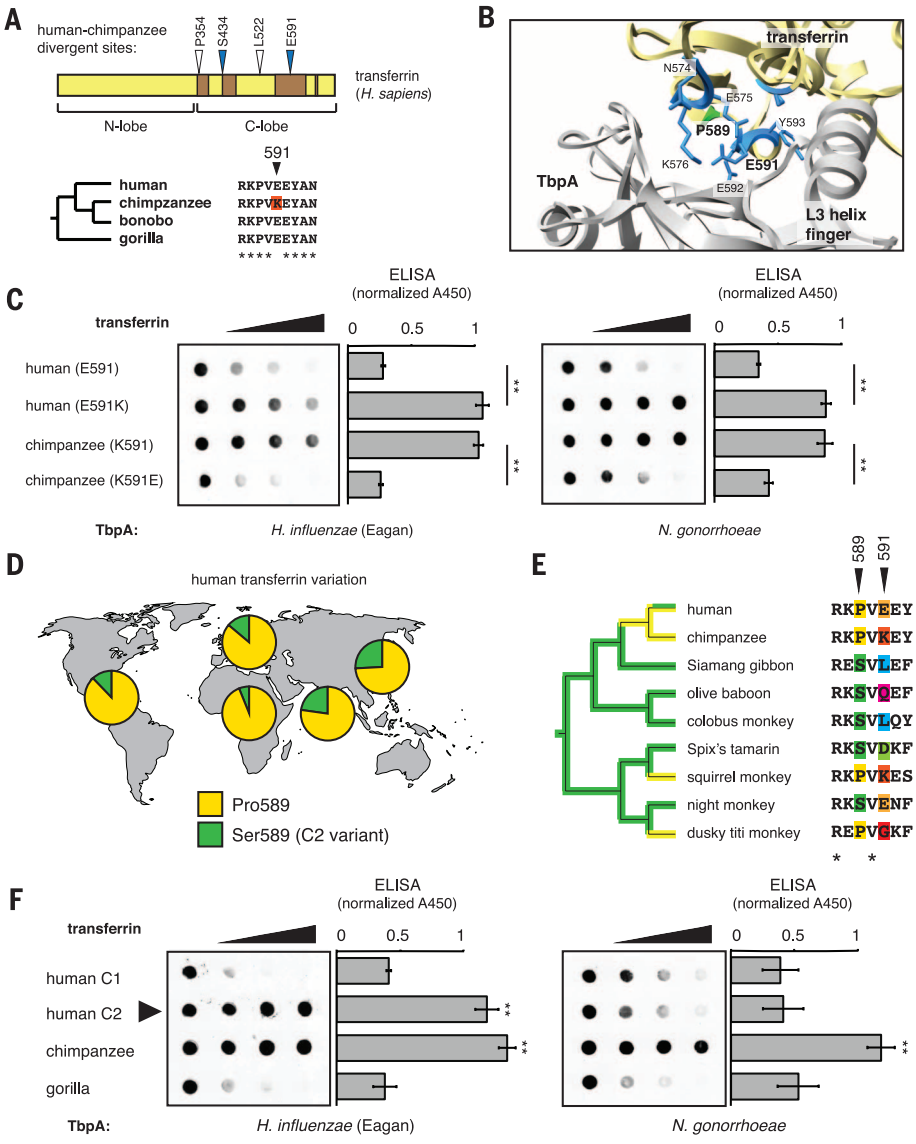


Fig. 2. Transferrin divergence in humans and chimpanzees impairs TbpA binding. (**A**) Schematic representation (top) showing divergent amino acid positions between human and chimpanzee transferrin. Blue arrows indicate amino acids that also display signatures of positive selection. Amino acid alignment (bottom) around position 591, highlighting the chimpanzee-specific E591 to K substitution. (**B**) Sites of positive selection in transferrin (blue) proximal to loop 3 of TbpA. Position 589, which is variable in human populations, is highlighted in green. (**C**) Competitive binding dot blots and ELISAs using recombinant human and chimpanzee transferrin, along with human E591K and chimpanzee K591E mutant proteins. Error bars represent SD of four independent experiments. $^{**}P < 0.01$. (**D**) Distribution of the transferrin C2 polymorphism (green) across human populations. (**E**) Primate phylogeny and amino acid alignment displaying toggling of transferrin position 589 across primates. Colors denote variable amino acids at each position. Variability at position 591 is also highlighted. (**F**) Competitive binding dot blot and ELISAs using the major transferrin variant (C1), the transferrin P589S variant (C2), chimpanzee, and gorilla transferrin. Error bars represent SD of four independent experiments. $^{**}P$ <0.01 relative to human (C1) transferrin.

A and B). Based on these findings, we speculate that there is an evolutionary trade-off between increased affinity to transferrin C2 and increased breadth of transferrin recognition. Regardless of whether such a trade-off constrains bacterial evolution, these findings reveal functionally distinct outcomes for transferrin recognition by TbpA among bacteria. Together our observations strongly suggest that both transferrin and TbpA have undergone repeated counteradaptations during the battle for iron over the course of primate evolution.

Observing functional consequences of both transferrin and TbpA evolution is consistent with the predictions of the "Red Queen" hypothesis, which posits that evolutionary arms races arise by recurrent episodes of positive selection between hosts and pathogens (*5*, *23*). To directly investigate whether TbpA has been subject to positive selection, we compared gene sequences from a large group of human-derived *Neisseria* and *H. influenzae* isolates (Fig. 3B and fig. S9). Horizontal gene transfer and recombination among bacterial strains notoriously compromise phylogenetic analyses of bacterial genes (*24*). Therefore, we relied on algorithms accounting for potential recombination break points and used combinations of structural insights and binding assays to substantiate predictions of positive selection (see the supplementary materials). We identified 10 sites among *Neisseria* and 9 among *Haemophilus* displaying strong signatures of positive selection, with one site shared between the two groups (Fig. 3, C and D, and tables S12 to S17). Nearly every site, though nonoverlapping between *Neisseria* and *Haemophilus*, lies in predicted extracellular loops of TbpA, which comprise the transferrin-binding interface (Fig. 3, C and D). These domains display marked variation even among closely related pathogen isolates, in contrast to a high degree of conservation in transmembrane domains (figs. S10 and S11).

To assess the functional implications of TbpA substitutions at rapidly evolving sites, we mutagenized *N. gonorrhoeae* TbpA at three positions under positive selection to corresponding amino acids present in *H. influenzae* (strain Eagan) TbpA (fig. S12). Of these three mutations, R365Q exhibited reduced binding to human transferrin (Fig. 3E). Thus, substitutions at rapidly evolving sites in transferrin, as well as TbpA, modulate interactions at this protein interface. By integrating phylogenetic analyses with high-quality structural data and experimental approaches, our results provide a high-resolution view of molecular genetic dynamics on both sides of host-pathogen evolutionary arms races.

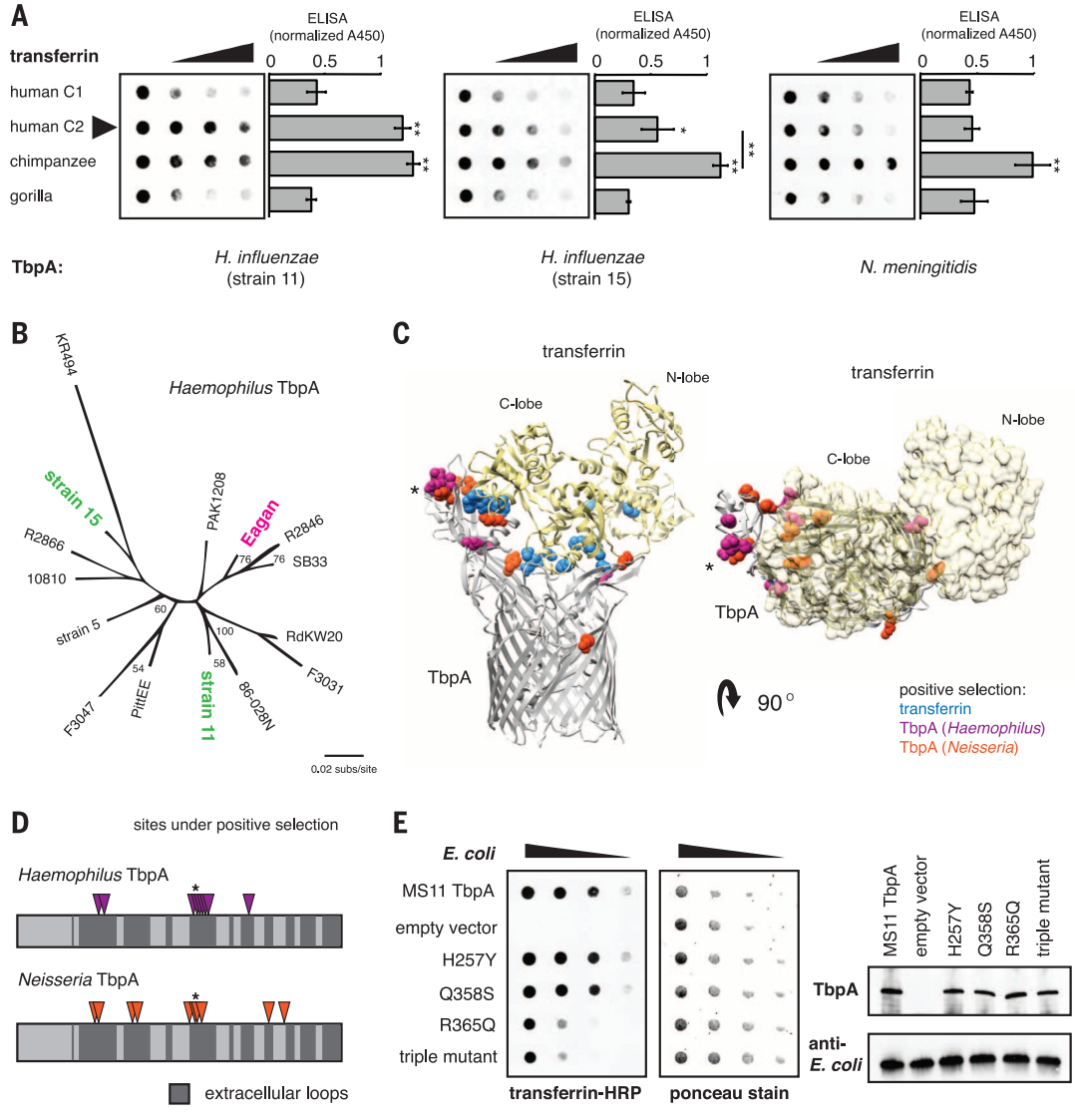


Fig. 3. Rapid evolution and functional variation of TbpA among human pathogens. (**A**) Competitive binding dot blots and ELISAs assessing TbpA binding to the human transferrin C2 variant. Error bars represent SD of four independent experiments. $^{**}P < 0.01$ relative to human (C1) transferrin, or C2 transferrin (vertical line). $^{*}P = 0.05$. (**B**) Gene tree of TbpA from *H. influenzae* isolates. TbpA from strain Eagan (magenta), as well as two divergent TbpA variants, strains 11 and 15 (green) are highlighted. (**C**) Structure of human transferrin (yellow) in complex with *N. meningitidis* TbpA (gray). Side chains of amino acids under positive selection in transferrin (blue), *H. influenzae* TbpA (purple), and *Neisseria* TbpA (orange) are denoted by colored spheres. * indicates position under positive selection in both *Haemophilus* and *Neisseria* TbpA. A "top" view of TbpA (right) shows exposure of rapidly evolving sites at the transferrin binding interface. (**D**) Schematic representation of sites among either *H. influenzae* or *Neisseria* TbpA showing evidence of positive selection, overlaid on TbpA from *N. meningitidis*. Arrows denote sites that pass multiple tests of positive selection, incorporating PhyML TbpA gene trees as well as phylogenies that account for recombination break points. Predicted extracellular loops are indicated in dark gray. * indicates a single amino acid position showing evidence of positive selection in both *Haemophilus* and *Neisseria*. (**E**) Indicated *E. coli* strains expressing mutations of TbpA were tested for interactions with human transferrin-HRP (left). A control blot was stained with ponceau (middle) as a loading control. Western blots using antibodies against *N. gonorrhoeae* TbpA or total *E. coli*. were performed with cell lysates from indicated strains (right).

In addition to the C1/C2 polymorphism in human transferrin, position 589 toggles exclusively between proline and serine across the primate lineage (Fig. 2E and fig. S13), a potential signature of antagonistic pleiotropy at a largely constrained position, as observed for other host-pathogen interfaces (*7*). Previous work has also implicated the C2 transferrin variant as a risk factor for disorders involving iron metabolism, including Alzheimer's disease; however, these associations remain controversial and appear dependent on the populations tested and interactions with other susceptibility loci (*25*, *26*). Our findings provide a functional basis for human transferrin variation and establish an important role for nutritional immunity in recent human evolution.

Although canonical innate immunity factors have been appreciated as nodes of host-virus evolution, our work demonstrates that nutritional immunity has played a fundamental role in the survival of primate populations challenged by bacterial pathogens. *H. influenzae* and *N. meningitidis* remain a major source of morbidity and mortality in regions where vaccine coverage is poor (*27*, *28*) and drug-resistant *N. gonorrhoeae* is developing into an urgent public health threat (*29*). By illuminating the battle for iron as a major driving force of host-pathogen evolution, from 40 million years of primate divergence to emerging human epidemics today, our studies reveal new reservoirs of genetic resistance to infectious diseases.

REFERENCES AND NOTES

1. E. D. Weinberg, *JAMA* **231**, 39–41 (1975).
2. J. E. Cassat, E. P. Skaar, *Cell Host Microbe* **13**, 509–519 (2013).
3. M. I. Hood, E. P. Skaar, *Nat. Rev. Microbiol.* **10**, 525–537 (2012).
4. I. Park *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 3149–3153 (1985).
5. M. D. Daugherty, H. S. Malik, *Annu. Rev. Genet.* **46**, 677–700 (2012).
6. A. Demogines, J. Abraham, H. Choe, M. Farzan, S. L. Sawyer, *PLOS Biol.* **11**, e1001571 (2013).
7. N. C. Elde, S. J. Child, A. P. Geballe, H. S. Malik, *Nature* **457**, 485–489 (2009).
8. M. R. Patel, Y.-M. Loo, S. M. Horner, M. Gale Jr., H. S. Malik, *PLOS Biol.* **10**, e1001282 (2012).
9. J. T. Kaelber *et al.*, *PLOS Pathog.* **8**, e1002666 (2012).
10. N. Noinaj *et al.*, *Nature* **483**, 53–58 (2012).
11. A. B. Schryvers, L. J. Morris, *Mol. Microbiol.* **2**, 281–288 (1988).
12. A. B. Schryvers, S. Gray-Owen, *J. Infect. Dis.* **165** (Suppl 1), S103–S104 (1992).
13. J. Alcantara, R. H. Yu, A. B. Schryvers, *Mol. Microbiol.* **8**, 1135–1143 (1993).
14. A. B. Schryvers, G. C. Gonzalez, *Infect. Immun.* **57**, 2425–2429 (1989).
15. M. L. Zarantonelli *et al.*, *Infect. Immun.* **75**, 5609–5614 (2007).
16. C. Calmettes, J. Alcantara, R.-H. Yu, A. B. Schryvers, T. F. Moraes, *Nat. Struct. Mol. Biol.* **19**, 358–360 (2012).
17. C. N. Cornelissen, G. D. Biswas, P. F. Sparling, *J. Bacteriol.* **175**, 2448–2450 (1993).
18. A. B. Schryvers, B. C. Lee, *Can. J. Microbiol.* **35**, 409–415 (1989).
19. S. D. Gray-Owen, A. B. Schryvers, *Microb. Pathog.* **14**, 389–398 (1993).
20. B. C. Lee, A. B. Schryvers, *Mol. Microbiol.* **2**, 827–829 (1988).
21. P. Kühnl, W. Spielmann, *Hum. Genet.* **43**, 91–95 (1978).
22. P. Zatta *et al.*, *Biochimica et Biophysica Acta (BBA) Molecular Basis of Disease* **1741**, 264–270 (2005).
23. M. Emerman, H. S. Malik, *PLOS Biol.* **8**, e1000301 (2010).
24. B. J. Shapiro, L. A. David, J. Friedman, E. J. Alm, . *Trends Microbiol.* **17**, 196–204 (2009).
25. K. Namekata *et al.*, *Hum. Genet.* **101**, 126–129 (1997).
26. Y. Wang *et al.*, *Can. J. Neurol. Sci.* **40**, 691–697 (2013).
27. K. L. O'Brien *et al.*, *Lancet* **374**, 893–902 (2009).
28. S. A. Halperin *et al.*, *Vaccine* **30** (suppl. 2), B26–B36 (2012).
29. T. Frieden, *Antibiotic Resistance Threats in the United States 2013* (Centers for Disease Control and Prevention, U.S. Department of Health and Human Services, 2013).

ACKNOWLEDGMENTS

We thank M. Mulvey and members of the Mulvey laboratory for assistance with bacterial genetics and helpful discussions; V. Chandrasekaran and W. Sundquist for assistance with baculoviral protein purification; J. Kaplan and D. Ward for helpful advice and discussions; J. St. Geme III and E. Porsch for *Haemophilus* strains and helpful discussions; M. So for providing *Neisseria* DNA and strains; S. Wong for providing rhesus macaque liver samples; Z. Kronenberg for assistance with human population genetics; and C. Cornelissen for the gift of the TbpA antibody as well as helpful discussions. J. Iwasa created the molecular animation of transferrin and TbpA evolution. We are grateful to N. Barber, H. Malik, S. Sawyer, N. Phadnis, and members of the Elde laboratory for comments on the manuscript. GenBank accession numbers of transferrin and TbpA variants are KM972645 to KM972665. Also see the supplementary materials. This work is supported by awards from the Pew Charitable Trusts and NIH to N.C.E. (GM090042) and M.F.B. (1F32GM108288). N.C.E. is a Pew Scholar in the Biomedical Sciences and Mario R. Capecchi Endowed Chair in Genetics.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/346/6215/1362/suppl/DC1
Materials and Methods
Figs. S1 to S13
Tables S1 to S18
Movie S1
References (*30*–*36*)

29 July 2014; accepted 14 November 2014
10.1126/science.1259329

POLITICAL SCIENCE

When contact changes minds: An experiment on transmission of support for gay equality

Michael J. LaCour[1] and Donald P. Green[2]

Can a single conversation change minds on divisive social issues, such as same-sex marriage? A randomized placebo-controlled trial assessed whether gay (n = 22) or straight (n = 19) messengers were effective at encouraging voters (n = 972) to support same-sex marriage and whether attitude change persisted and spread to others in voters' social networks. The results, measured by an unrelated panel survey, show that both gay and straight canvassers produced large effects initially, but only gay canvassers' effects persisted in 3-week, 6-week, and 9-month follow-ups. We also find strong evidence of within-household transmission of opinion change, but only in the wake of conversations with gay canvassers. Contact with gay canvassers further caused substantial change in the ratings of gay men and lesbians more generally. These large, persistent, and contagious effects were confirmed by a follow-up experiment. Contact with minorities coupled with discussion of issues pertinent to them is capable of producing a cascade of opinion change.

Foremost among theories of prejudice reduction (*1*) is the contact hypothesis (*2*), which contends that outgroup hostility diminishes when people from different groups interact with one another. Although contact is credited with reducing prejudice toward a wide array of outgroups (*3*), in practice it is often difficult to facilitate intergroup contact of sufficient duration to dispel negative stereotypes and build empathy. For this reason, research attention has recently focused on alternative interventions that may be deployed in a more compressed time frame. Examples include brief personal contact with outgroup members during the course of a conversation (*4*) and the "extended contact" that occurs when one learns that a close friend has experienced positive contact with an outgroup (*5*). The question is whether brief or indirect contact is sufficient to produce meaningful and enduring attitude change. Recent literature reviews have been tentative on this point, noting the lack of randomized experiments that track attitudes months after the intervention (*6*).

Our theoretical contribution is to introduce the distinction between active and passive contact, which are posited to produce different effects in the context of a brief intergroup encounter. Whereas passive contact involves personal exposure to an outgroup member (e.g., through collaborative activity), active contact involves, in addition, communication about an issue that divides the two groups (e.g., discussion of recent communal violence). The effects of active contact doubtless depend on whether the conversation is respectful or accusatory, but in principle, active contact has the potential to both reduce hostility toward outgroups and to change attitudes on divisive issues. Our empirical contribution is the first field-based experimental demonstration of persistent attitude change in the wake of active

[1]Department of Political Science, University of California, Los Angeles (UCLA), Los Angeles, CA, USA. [2]Department of Political Science, Columbia University, New York, NY, USA.

Downloaded from www.sciencemag.org on December 14, 2014

contact. The effects were substantively large among those who received the messages directly, and these effects diffused to other members of the receiver's household. Both direct and secondhand effects were estimated with a high degree of statistical precision, and findings were confirmed in a follow-up experiment.

Experimental design

Overview

Registered voters who had previously been enrolled in an Internet panel survey were contacted at their doorstep by canvassers. Random assignment determined whether contact was initiated by a gay or straight canvasser and whether the canvasser discussed the subject of same-sex marriage or recycling. Outcomes were assessed unobtrusively by online surveys conducted days, weeks, and months afterward.

Participants

Participants in the study were Southern California residents who (i) are registered to vote in precincts that supported a ballot measure banning same-sex marriage in 2008, (ii) reside at the same address as at least one other registered voter, and (iii) live in neighborhoods composed of detached dwellings (e.g., single-family homes). Using the publicly available California voter file as our sampling frame, we first recruited voters meeting the above criteria to participate in an online survey panel about politics. Second, in an effort to impanel multiple voters per household, individuals were offered additional financial incentives to refer their friends and family to participate in the survey panel. If the referred individual was a registered voter who met the above criteria, he or she was invited to participate in the study. Respondents received compensation for each successful referral. This recruitment procedure enables us to estimate how experimentally induced treatments diffuse through voters' social networks (*7*, *8*).

Random assignment

Households in which at least two registered voters completed the first wave of the online panel survey (29–30 May 2013) were randomly assigned to five experimental conditions. Simple random assignment occurred at the household level to facilitate the analysis of within-household spillovers. The first group was assigned to receive the same-sex marriage script from a gay canvasser. The second group was assigned to receive the same-sex marriage script from a straight canvasser. Groups three and four were encouraged to recycle household waste by gay or straight canvassers, respectively; however, canvassers did not reveal their sexual orientation when delivering the recycling script. The fifth group was a control group to which no canvassers were assigned. The resulting distribution of household assignments is described in table S3.

Canvassers and contact

Canvassers were recruited and trained by the Los Angeles LGBT Center, our nongovernmental organization partner. When approaching each targeted address, canvassers were instructed to (i) administer the assigned script to the first voter who answered the door, (ii) speak to only one voter per household, and (iii) confirm his or her name. The purpose of speaking with and identifying a single voter in each household is to facilitate the estimation of spillover effects due to within-household conversations, as explained below. Apart from a single canvassing conversation on 1 June 2013, neither the canvassers nor the center had further interaction with voters.

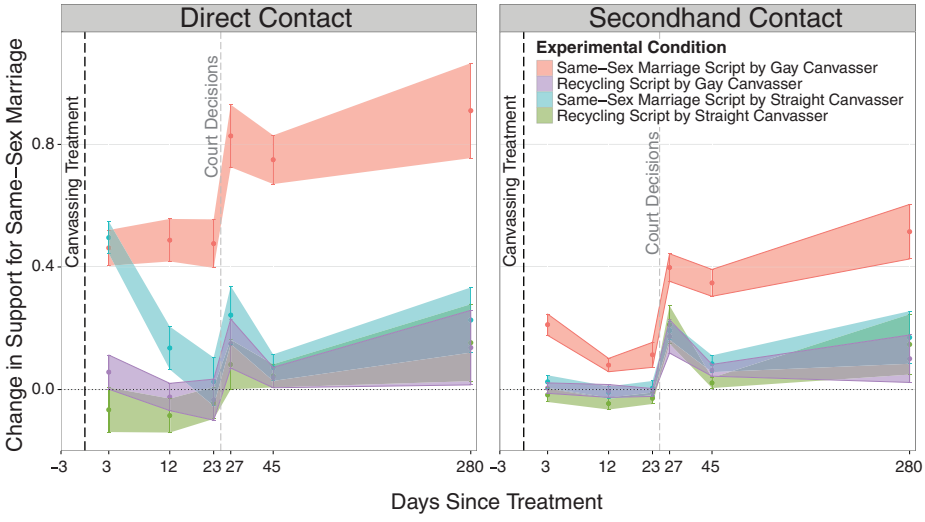


Fig. 1. Direct and secondhand effects on support for same-sex marriage, by assigned message and messenger, and time since treatment. The first vertical dashed line represents the canvassing intervention, which was administered between Internet survey waves 1 and 2. The second vertical dashed line represents the U.S. Supreme Court decisions striking down California's ban on same-sex marriage. The *y* axis is opinion change between the baseline survey and subsequent survey waves, with higher scores indicating more support for same-sex marriage. Points represent mean values, and bars display 95% bootstrap confidence intervals.

A survey administered to canvassers measured demographic attributes, canvassing experience, and sexual orientation. Table S1 shows that gay (n = 22) and straight (n = 19) canvassers are similar on dimensions other than sexual orientation. This characterization was confirmed by independent coders, who viewed videotaped recordings of the canvassers and rated their personal attributes and communication style (table S2).

Messages

Canvassers were coached to be polite and respectful at all times, to listen attentively to voters when discussing either same-sex marriage or recycling, and to refrain from arguing with voters. Talking points for the same-sex marriage and recycling scripts are presented in fig. S1. The same-sex marriage script invited voters to share their experiences with marriage. This script was the same for gay and straight canvassers, with one important exception. After establishing rapport with the voter, midway through the conversation gay canvassers revealed that they are gay or lesbian and that they would like to get married but that the law prohibits same-sex marriage. Straight canvassers instead described how their child, friend, or relative would like to get married but that the law prohibits same-sex marriage. Voters were asked to share their thoughts on this dilemma. These doorstep conversations lasted on average 22 min.

Outcome measurement

Before canvassers went into the field, the study first gathered baseline positions on a range of political attitudes for all participants, using an ostensibly unrelated online survey. The survey included 50 questions, as described in the supplementary materials. The two questions concerning same-sex marriage and feelings about gay people were buried amid a large number of items on unrelated topics, so that respondents would not suspect any connection between the survey and the canvassing visit. The panel survey included the two outcome measures in every wave, but the overall content of the survey was kept fresh by rotating new questions in each wave. Another important feature of the design concerns blinding of canvassers and survey respondents. Voters were unaware that the online survey was related to the canvassing effort, and canvassers had no knowledge that voters were participants in an online survey. Finally, to prevent housemates from completing each other's surveys, extensive precautions were taken, including issuing distinct personal login instructions to take differently named surveys, tracking of distinct IP addresses and session cookies, and sending invitations to take the survey during working hours (91% were in fact completed at work), when housemates would be less likely to be together.

Verifying design assumptions

Statistical tests presented in tables S4 to S7 confirm that (i) the assigned experimental groups have statistically indistinguishable background attributes and baseline opinions about same-sex

marriage; (ii) canvassers contacted a similar proportion of their intended targets, regardless of whether the message to be delivered concerned same-sex marriage or recycling; and (iii) survey panel attrition occurred at a similar rate across the five experimental conditions.

Identifying causal effects

Because voters who answered the door were randomly presented with either a treatment (same-sex marriage) or placebo (recycling) script, their average treatment effect may be estimated by comparing average outcomes among those who received the same-sex marriage script to outcomes among those who received the recycling script. All people who answered the door to canvassers were coded as receiving the script, regardless of the length of the ensuing conversation. In 11 cases, canvassers were told to go away before completing their script; we code these attempts as successful contacts for the purposes of (conservatively) estimating the average treatment effect among door-answerers (*9*), as explained in the supplementary materials. Treatment versus placebo comparisons may be further subdivided according to whether the canvasser was gay or straight, as every canvasser delivered both the same-sex marriage and recycling scripts in random sequence. In order to estimate the spillover effect of communication within the household, we compared the housemates of those who received the same-sex marriage script to the housemates of those who received the recycling script.

Results

Direct effects of conversations with canvassers

The left panel of Fig. 1 tracks responses to the same-sex marriage question among voters who conversed directly with canvassers in the four experimental conditions. The question reads: "Do you favor or oppose allowing gays and lesbians to marry legally?" Response options ranged from strongly oppose (1) to strongly favor (5). Outcomes displayed in Fig. 1 are the change in views toward same-sex marriage between the baseline survey and each post-treatment survey. The six post-treatment surveys occurred 3, 12, 23, 27, 45, and 280 days after canvassing took place.

Before the canvassing intervention, all four randomly assigned groups expressed similar views toward same-sex marriage. Three days after treatment, however, the groups diverged markedly. Figure 1 indicates that those receiving the recycling message became no more supportive of same-sex marriage. In contrast, those receiving the same-sex marriage message from gay canvassers became 0.46 scale points more supportive of same-sex marriage. Those who spoke with straight canvassers about same-sex marriage became 0.50 scale points more supportive. As Table 1 indicates, the latter two message effects are each statistically significant at $P < 0.0001$, using a two-tailed test.

The treatment effects for people who conversed with gay or straight canvassers are not only statistically significant; they are substantively large. Pooling over both types of canvassers, the estimated effect of the message is approximately 0.48. To put this estimate in perspective, note that when the same question was asked in the 2012 Cooperative Campaign Analysis Project (CCAP), an online survey of the national adult population, the average response of residents of Nebraska and Ohio (3.14) was roughly the same as the average response in the treatment group before visits from canvassers. The same-sex marriage script bumped the treatment group up by an average of 0.48 points to roughly the same mean as respondents from Connecticut and Massachusetts (3.61). On the issue of same-sex marriage, the canvassing treatment in effect transformed Midwesterners into New Englanders.

Tracing direct opinion change effects over time

The preceding discussion suggests that canvassing had a substantial effect on policy views expressed less than a week after the intervention. Although the treatment scripts employed by gay

Table 1. Experimental results. Support for same-sex marriage in study 1, by assigned message and messenger, form of contact, and time elapsed since treatment. Entries in plain type are the average changes in support for same-sex marriage, where changes are the difference from pretreatment survey responses (wave 1). Entries in boldface indicate the estimated average treatment effect of message for each messenger. The same-sex marriage question drawn from (*12*) reads: "Do you favor or oppose allowing gays and lesbians to marry legally?" Response options ranged from strongly oppose (1) to strongly favor (5), forming a five-point scale. "Treatment assigned" refers to the randomly assigned message and messenger pairings. "Contact" refers to whether the respondent spoke directly with a canvasser ("direct") or was the housemate of someone who spoke directly with canvassers ("secondhand"). *n* indicates the number of voters who answered each online survey, 5 days and 9 months post-treatment. See table S6 for the number of responses by condition and wave. Standard errors (shown in parentheses) are clustered at the household level for secondhand contact.

		Direct contact		Secondhand contact	
Treatment assigned		**Time since treatment**			
Message	Messenger	5 days	9 months	5 days	9 months
Same-sex marriage	Gay	0.46 (0.05)	0.91 (0.14)	0.21 (0.04)	0.52 (0.09)
Recycling	Gay	0.06 (0.05)	0.14 (0.11)	0.00 (0.02)	0.10 (0.07)
Treatment effect for gay messengers		**0.41 (0.07)**	**0.77 (0.18)**	**0.21 (0.04)**	**0.41 (0.11)**
n		242	184	437	324
Same-sex marriage	Straight	0.50 (0.05)	0.23 (0.10)	0.03 (0.02)	0.17 (0.08)
Recycling	Straight	−0.07 (0.07)	0.15 (0.11)	−0.02 (0.02)	0.15 (0.09)
Treatment effect for straight messengers		**0.56 (0.08)**	**0.07 (0.15)**	**0.04 (0.03)**	**0.02 (0.12)**
n		225	182	371	282

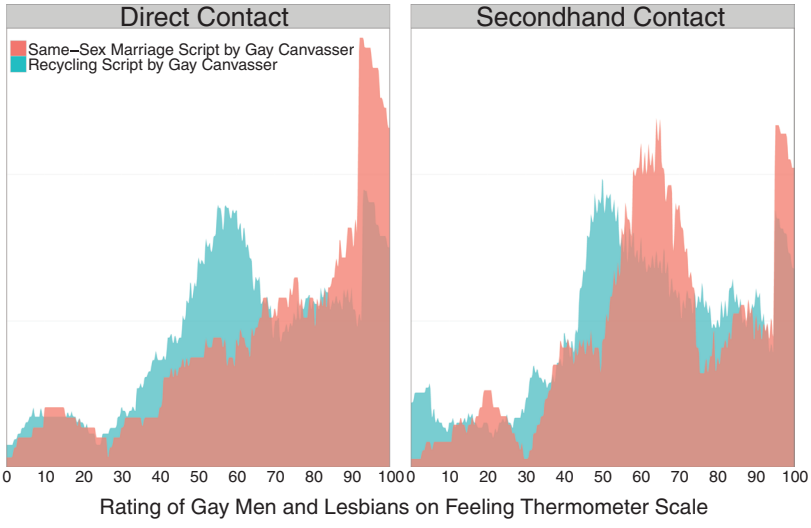


Fig. 2. Direct and secondhand contact effects on evaluations of gay men and lesbians 9 months after canvassing. The figure displays rectangular kernel density estimates of ratings of gay men and lesbians on the feeling thermometer scale 0 (unfavorable) to 100 (favorable) in study 1, among those contacted by gay canvassers. As shown in table S13, the large, persistent, and contagious effects of contact with gay canvassers on ratings of gay men and lesbians were replicated in study 2.

and straight canvassers were slightly different, they proved to be equally effective in changing minds in the short run. How long-lasting were these shifts in opinion? Those who discussed same-sex marriage with straight canvassers quickly reverted to their pretreatment baseline opinions, and 90% of the initial treatment effect dissipated a month after the conversation with canvassers. Some of the initial effect was revived in the immediate aftermath of the landmark Supreme Court decisions of 26 June 2013, which effectively legalized same-sex marriage in California. However, this surge in support, too, subsided within a few days. In contrast, among those who discussed same-sex marriage with gay canvassers, the treatment effects remained unabated after 1 month and strengthened markedly after the Supreme Court decisions. Nine months after their initial conversations with gay canvassers, those who conversed about same-sex marriage were 0.77 scale points more supportive of this policy than those who discussed recycling ($P < 0.0001$). Again, to put this treatment effect into perspective, note that in the 2012 CCAP survey, 0.78 scale points separate the states of Georgia and Massachusetts.

Spillover effects

How much of the canvassing effect was transmitted to housemates of the voters who answered the door? Figure 1 (right panel) presents housemates' average support for same-sex marriage, classified according to the experimental condition to which their household was assigned. Although housemates did not receive the gay equality message directly, our evidence strongly suggests that they were influenced by secondhand exposure to the treatment. Housemates of those who received the placebo treatments expressed opinions about same-sex marriage that were essentially unchanged over time. Those whose housemates conversed with gay canvassers about same-sex marriage became 0.21 scale points more supportive of same-sex marriage when surveyed 3 days later (two-tailed $P < 0.0001$). No apparent change occurred among respondents whose housemates conversed with straight canvassers about same-sex marriage. Although the spillover effect we observed after 3 days subsided 3 weeks later, it was rekindled after the Supreme Court decisions. Thereafter, all groups became more supportive of same-sex marriage initially, but those whose housemates conversed with gay canvassers about same-sex marriage moved significantly farther than their counterparts in other groups. Six months later, cohabitants of those who spoke with gay canvassers about same-sex marriage were 0.41 scale points more supportive of same-sex marriage than cohabitants of those who spoke with gay canvassers about recycling. The two-tailed P value of the estimated long-term spillover effect is just 0.0002.

A follow-up experiment to verify key findings

In August of 2013, we conducted a second study to verify the three key findings from study 1: large, persistent, and contagious effects of conversations with gay canvassers discussing same-sex marriage. All three findings were confirmed with similar effect sizes and P values of less than 0.001, as described in tables S10 to S13.

Further evidence suggesting causal mechanisms

Our results contribute new evidence to the literature on opinion change and prejudice reduction. Face-to-face appeals on behalf of same-sex marriage led people who conversed with canvassers to express substantially higher levels of support for this proposal when their views were measured a few days later. However, the script itself recedes in importance as we trace the effects over time or through voters' social networks. Gay and straight canvassers produced distinctive patterns of attitude change. Both gay and straight canvassers produced large effects of similar magnitude initially, but only gay canvassers' effects persisted in 3-week, 6-week, and 9-month follow-ups.

Three additional statistical results provide hints about the psychological process by which attitude change occurred in the wake of contact with gay canvassers. The first is that overall evaluations of gay men and lesbians became significantly more positive. Figure 2 shows how voters rated "gay men and lesbians" on a scale ranging from 0 (cold) to 100 (warm) 9 months after canvassing took place. Voters who spoke with gay canvassers about recycling gave ratings that were on average 4.5 points higher than their baseline rating before the intervention. In contrast, voters who spoke with gay canvassers about same-sex marriage offered average ratings that were 15.1 points higher than the scores they registered at baseline. This effect ($15.1 - 4.46 = 10.64$) is significant at $P < 0.001$ (table S8). A highly significant spillover effect ($10.25 - 3.23 = 7.02$) was also apparent among the housemates of those who spoke with gay canvassers about same-sex marriage (P value < 0.0001, table S9). The same-sex marriage script produced no such long-term direct or spillover effects when delivered by straight canvassers.

The preceding fact suggests that contact changes attitudes, but what evidence distinguishes active from passive contact? The second fact comes from a follow-up experiment conducted on a fresh set of voters drawn at random from the same pool of survey respondents. In this study, gay canvassers discussed either the subject of abortion or a placebo topic (recycling). Space constraints prevent us from describing the study in detail, but the key finding is that lengthy conversations with canvassers who revealed themselves to be gay while discussing abortion rights had no effect on people's opinions about same-sex marriage or gay people more generally. Evidently, the profound attitude change described in Fig. 1 hinges on active contact: discussing gay equality with an openly gay person.

A further indication of the distinctive impact of contact with gay canvassers is that their treatment message resonated throughout the household, as would be predicted by theories of extended contact (*10*). Only gay canvassers produced large and statistically significant secondhand effects on the housemates of those who came to the door. It may be that a conversation with a gay canvasser about same-sex marriage was more likely to be recounted to a housemate. Or it may be that the attitude change brought about by gay canvassers was more deeply felt by voters who participated in the conversation, which in turn made them more effective spokespersons for their newfound view. Interestingly, the secondhand effects of contact with gay canvassers were amplified after the Supreme Court decisions, suggesting that the extensive news coverage (*11*) surrounding this change in policy sparked new conversations and concomitant attitude change within treated households.

Our experimental results demonstrate that active contact is capable of producing a cascade of enduring opinion change. Further research is needed to assess the extent to which the strength, diffusion, and persistence of active contact's effects depend on how groups come together, the salience of their identities, the issues they discuss, and the manner in which deliberation takes place.

REFERENCES AND NOTES

1. E. L. Paluck, D. P. Green, *Annu. Rev. Psychol.* **60**, 339–367 (2009).
2. G. W. Allport, *The Nature of Prejudice* (Doubleday, New York, 1954).
3. T. F. Pettigrew, L. R. Tropp, *J. Pers. Soc. Psychol.* **90**, 751–783 (2006).
4. N. C. Scarberry, C. D. Ratcliff, C. G. Lord, D. L. Lanicek, D. M. Desforges, *Personality Social Psychol.* **23**, 1291–1299 (1997).
5. S. C. Wright, A. Aron, T. McLaughlin-Volpe, S. A. Ropp, *J. Pers. Soc. Psychol.* **73**, 73–90 (1997).
6. M. Hewstone, H. Swart, *Br. J. Soc. Psychol.* **50**, 374–386 (2011).
7. R. M. Bond *et al*., *Nature* **489**, 295–298 (2012).
8. K. J. Gile, M. S. Handcock, *Sociol. Methodol.* **40**, 285–327 (2010).
9. A. S. Gerber, D. P. Green, *Field Experiments: Design Analysis and Interpretation* (W.W. Norton, New York, 2012).
10. K. Lewin, in *Readings in Social Psychology* (Henry Holt and Company, New York, 1952).
11. P. Hitlin, M. Jurowitz, A. Mitchell, "News coverage conveys strong momentum for same-sex marriage" (Pew Research Center, Washington, DC, 2013.
12. S. Jackman, J. Sides, M. Tesler, L. Vavreck, *Common Content Codebook, Cooperative Campaign Analysis Project* (YouGov, Palo Alto, CA, 2012).

ACKNOWLEDGMENTS

M.J.L. received research funding from the Williams Institute, Ford Foundation, and Evelyn and Walter Haas, Jr., Fund, which bear no responsibility for the findings or interpretations given here. Replication data and accompanying R code are available at OpenICPSR (http://doi.org/10.3886/E19260V1). Neither of the authors is affiliated with the Los Angeles LGBT Center or received compensation for this research. The experiments reported here were registered at the Experiments in Governance and Politics site before the launch of each study. This research was approved by the UCLA institutional review board, case no. 12-001750.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/346/6215/1366/suppl/DC1
Materials and Methods
Fig. S1
Tables S1 to S14
References (*13*–*25*)

16 May 2014; accepted 7 November 2014
10.1126/science.1256151

SYSTEMS BIOLOGY

Accurate information transmission through dynamic biochemical signaling networks

Jangir Selimkhanov,[1*] Brooks Taylor,[1*] Jason Yao,[2] Anna Pilko,[2] John Albeck,[3] Alexander Hoffmann,[4,5] Lev Tsimring,[4,6] Roy Wollman[2,4,7†]

Stochasticity inherent to biochemical reactions (intrinsic noise) and variability in cellular states (extrinsic noise) degrade information transmitted through signaling networks. We analyzed the ability of temporal signal modulation—that is, dynamics—to reduce noise-induced information loss. In the extracellular signal–regulated kinase (ERK), calcium (Ca^{2+}), and nuclear factor kappa-B (NF-κB) pathways, response dynamics resulted in significantly greater information transmission capacities compared to nondynamic responses. Theoretical analysis demonstrated that signaling dynamics has a key role in overcoming extrinsic noise. Experimental measurements of information transmission in the ERK network under varying signal-to-noise levels confirmed our predictions and showed that signaling dynamics mitigate, and can potentially eliminate, extrinsic noise–induced information loss. By curbing the information-degrading effects of cell-to-cell variability, dynamic responses substantially increase the accuracy of biochemical signaling networks.

The role of biological signaling networks is to reliably transmit specific information about the extracellular environment to downstream effectors, allowing the cell to adjust its physiological state to changing conditions. The stochasticity of molecular interactions that underlies various forms of "noise" in biological systems can interfere with signal transduction and degrade the transmitted information. How signaling networks perform their core functions in the presence of noise is a fundamental question. Information-theoretic approaches allow estimation of the maximal possible information transmission capacity of noisy biochemical networks (*1*–*11*). Previous applications of such methods to the analysis of signaling networks suggested that as a result of noise, cells lose most of the information about the concentration of ligands (*12*–*14*). Thus far, the information-theoretic analyses of signaling networks have been based on scalar measurements performed at a single time point. However, the information on activating ligands is often encoded using a dynamic signal represented by a multivariate vector that contains a single cell's response at multiple time points (*15*–*18*).

[1]Department of Bioengineering, University of California–San Diego, La Jolla, CA 92093, USA. [2]Department of Chemistry and Biochemistry, University of California–San Diego, La Jolla, CA 92093, USA. [3]Department of Molecular and Cellular Biology, University of California–Davis, Davis 95616, USA. [4]San Diego Center for Systems Biology, La Jolla, CA 92093, USA. [5]Institute for Quantitative and Computational Biosciences and Department of Microbiology, Immunology, and Molecular Genetics, University of California–Los Angeles, Los Angeles, CA 90025, USA. [6]BioCircuits Institute, University of California–San Diego, La Jolla, CA 92093, USA. [7]Cell and Developmental Biology Section, Division of Biological Sciences, University of California–San Diego, La Jolla, CA 92093, USA.
*These authors contributed equally to this work. †Corresponding author. E-mail: rwollman@ucsd.edu

To test the hypothesis that dynamic responses contain more information than static responses, we performed single live-cell measurements of three key signaling pathways (Fig. 1): extracellular signal–regulated kinase (ERK), calcium (Ca^{2+}), and nuclear factor kappa-B (NF-κB) [supplementary materials (SM) section 1.1]. Fully automated computational image analysis (SM section 1.2) allowed us to measure the response of 910,121 individual live cells (figs. S1 to S6 and tables S1 to S3). The large sample size was instrumental for analyzing high-dimensional multivariate responses. In all three pathways, there was substantial variability within the dynamic (Fig. 1, C to E) and nondynamic (Fig. 1F) single-cell responses across multiple concentrations of activating ligands.

To analyze the implications of noise on information loss, we used an information-theoretic approach to calculate the information transmission capacity of a dynamic signaling network. The information transmission capacity [also referred to as channel capacity (*19*)] is measured as the maximal possible mutual information between the measured response and the activating ligand concentration. To calculate the mutual information between a dynamic response (a vector) and the ligand concentration (a scalar), we expanded on a previously described algorithm (*14*). The algorithm uses continuous multidimensional response data and a k–nearest-neighbor approach to estimate the conditional probability density for each cell's response (SM section 2). We thus estimated the information transmission capacity of the dynamic response and of several types of static responses. For all single–time point static scalar responses, we found transmission capacity (<1 bit) (*12*, *13*) (Fig. 2A). However, across all three signaling pathways, the dynamic response had significantly higher information transmission capacity than several scalar responses previously described (*20*, *21*) (Fig. 2, B and C, Student's t test, $P < 0.05$ for all comparisons, table S6). These estimates should be considered as lower bounds because they do not exclude variability resulting from experimental imperfections.

To elucidate the origins of the enhanced information transmission capacity of dynamic signaling responses, we developed a mathematical theory using information-theoretic formalism (SM section 3). The theory explicitly accounts for the information-degrading effects of intrinsic and extrinsic noise sources in the context of multivariate responses. Intrinsic noise adds to uncertainty in all dimensions (i.e., time points) independently from one another. In contrast, the extrinsic variability in cellular states produces fluctuations that are constrained by the signaling network that generates the dynamics. Therefore, the fluctuating components generated by extrinsic noise at different time points are deterministically dependent on one another. As a result, intrinsic and extrinsic noise sources have different effects on the information transmission capacity of multivariate responses. In the case of purely intrinsic noise, additional measurements increase the information logarithmically because of simple ensemble averaging (*12*). In the case of purely extrinsic noise, a sufficient number of dynamical measurements can provide complete information about the a priori uncertain internal state of the cell and therefore lead to a substantial gain in the information about the activating ligand (Fig. 3A).

To test our analytical prediction that the multivariate dynamic response can completely eliminate the information loss that results from introduction of extrinsic noise (SM section 3), we used computer simulations of ERK responses based on a published kinetic model (SM section 4.1) (*22*) (fig. S15). We generated sets of simulated ERK activity trajectories in response to an increasing number of ligand concentrations. We varied model values for ERK and mitogen-activated protein kinase kinase (MEK) according to a uniform distribution (±20% mean value) to mimic extrinsic noise and measured the information transmission capacity. Our analysis supported the analytical prediction and showed that whereas the univariate response, based on maximal ERK dynamics, had limited information transmission capacity, the dynamic multivariate response can transmit complete information about ligand concentration (Fig. 3B). An intuitive demonstration for the limitation of univariate response and the ability of multivariate response to overcome extrinsic noise is shown in Fig. 3, C and D. Superficially, the trajectories of two populations of simulated responses of ERK activity to two input concentrations of epidermal growth factor (EGF) appear overlapping (Fig. 3C), but in fact, they are completely separable when considering joint distributions (*23*). Plotting the distribution of ERK activity at t = 9 and t = 24 min on a two-dimensional (2D) plane (Fig. 3D) shows that the responses to a single varied parameter input lies on a one-dimensional (1D)

Downloaded from www.sciencemag.org on December 14, 2014

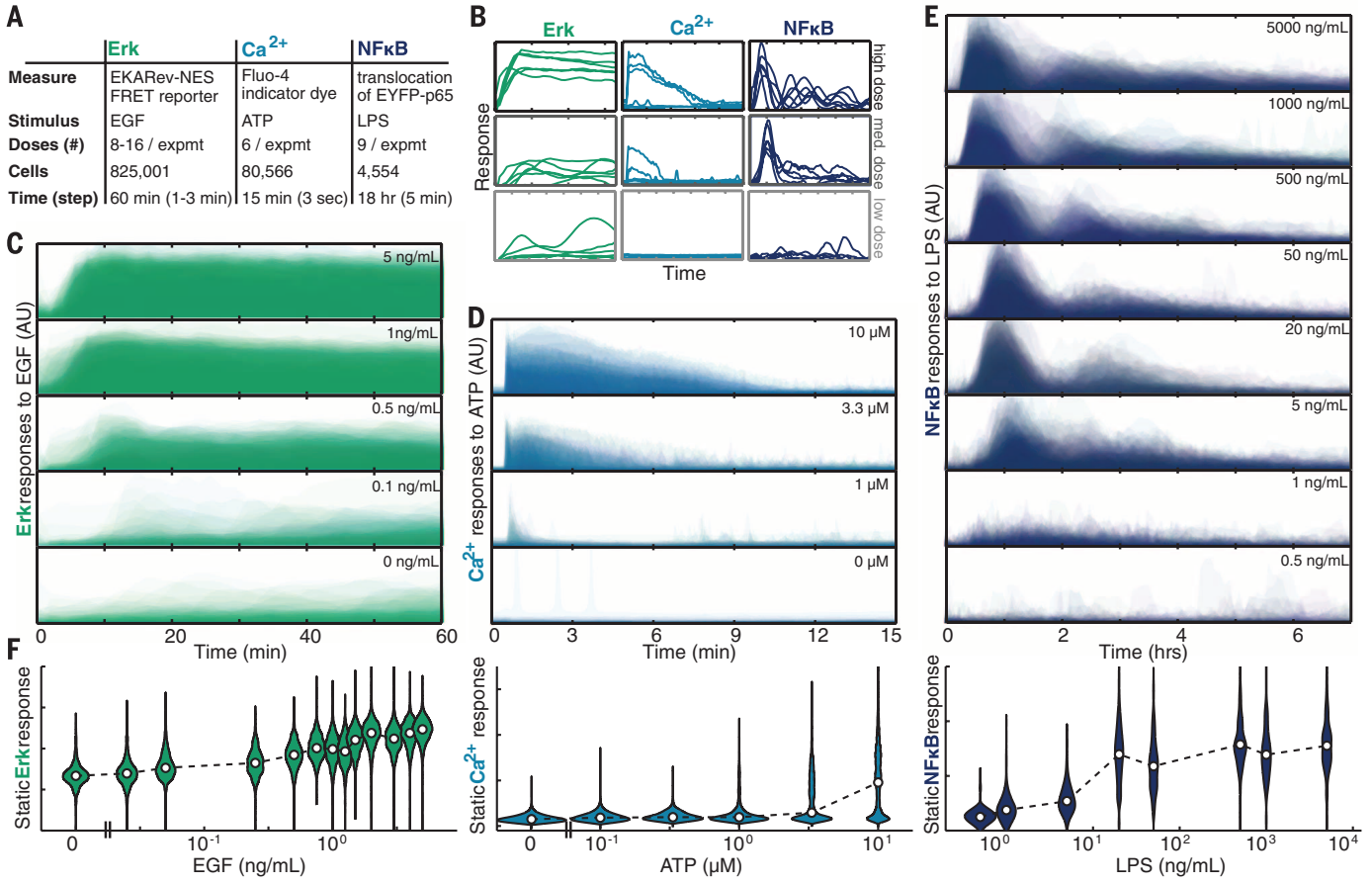


Fig. 1. Single-cell measurement of the dynamic response of ERK, Ca^{2+}, and NF-κB. (**A**) Overview of single-cell data analyzed in this work. (**B**) Examples of single-cell response dynamic trajectories. (**C** to **E**) Temporal histograms of several representative dosages for ERK (C), Ca^{2+} (D), and NF-κB (E). Color intensity reflects the probability density of a cellular response magnitude at each time point. *Y* axis in (B) to (E) is the same for each pathway and is in arbitrary units (AU), representing the Förster resonance energy transfer (FRET) to cyan fluorescent protein (CFP) ratio reported by the EKARev ERK biosensor (C), intensity of Ca^{2+} indicator dye Fluo-4 (D), and ratio of nuclear to cytoplasmic localization of an enhanced yellow fluorescent protein (EYFP)–p65 reporter (E). (**F**) Violin plot of the maximally separable static response in the three signaling pathways. Shape width shows response distribution (areas are equal), and point is the median response in each condition. EGF, epidermal growth factor; ATP, adenosine triphosphate; LPS, lipopolysaccharide.

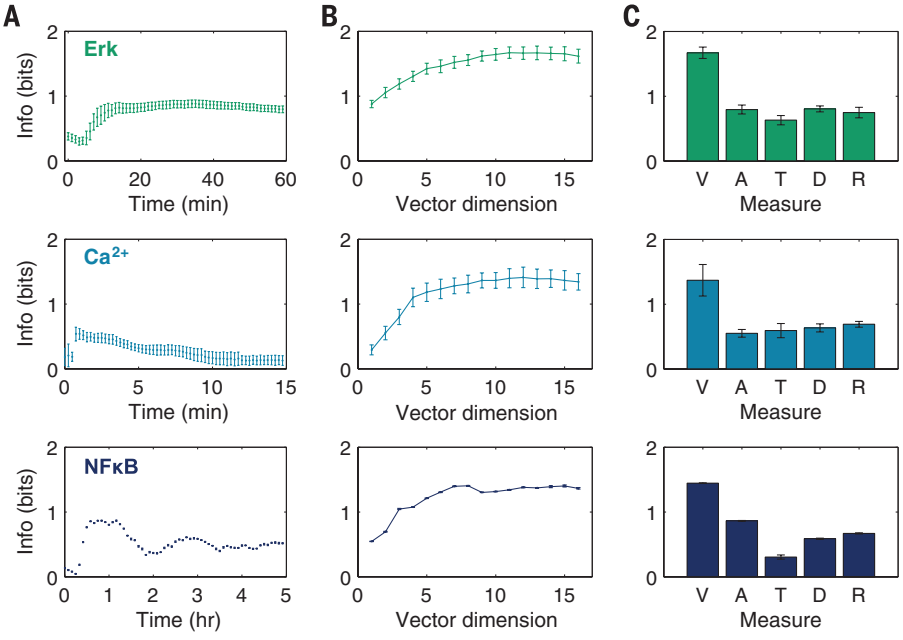


Fig. 2. Information transmission capacity of static and dynamic ERK, Ca^{2+}, and NF-κB responses. (**A**) Information transmission capacity calculated from static scalar response distribution based on single–time point measurements. (**B**) Information transmission capacity calculated from multivariate dynamic responses as a function of the dimension of the multivariate vector. The multivariate vector was subsampled using a uniform grid centered on the middle time point (fig. S19). (**C**) Comparison of the multivariate vector (V) measurement to the following scalar responses: maximum response amplitude (A), maximum response time (T), maximal rate of response (D), ratio of maximum response amplitude to initial response amplitude (R). Error bars are SEMs from six biological replicates for ERK and four for Ca^{2+}, and SDs from five jackknife iterations for NF-κB (tables S1 to S3). The multivariate vector information transfer was significantly greater than all scalar measures ($P < 0.05$, Student's *t* test, table S6).

curve within a 2D space. The two 1D manifolds for different inputs are completely separated from each other (inset), but overlap considerably in any 1D projection. This simple example demonstrates how the extrinsic variability of a single parameter can in principle be completely eliminated with measurements from only two time points.

The accuracy of a response can be characterized by its signal-to-noise ratio (SNR). The mutual information and the system's SNR are related; however, this relationship is strongly affected by the noise properties (intrinsic versus extrinsic) and the type of the response. Our analytical theory predicts a different relationship between mutual information and SNR for three different types of responses: (i) scalar responses that do not distinguish between intrinsic and extrinsic noise; (ii) multivariate responses without any dynamic component (redundant measurements) that can only reduce intrinsic noise; and (iii) dynamic responses that combine the benefits of redundant measurements with efficient mitigation of extrinsic variability. We varied the SNR in the ERK network by partial inhibition of the ERK kinase MEK with six different dosages of the inhibitor U0126. At each MEK inhibition level, we measured ERK response to eight EGF levels. A total of 48 conditions were measured in four biological replicates (fig. S23). At each MEK inhibitor level, we calculated the mutual information and the SNR from single-cell responses (SM section 4.3.1). In total, Fig. 4 contains 535,107 cell responses (tables S4 and S5). As expected, for a scalar response, the formula relating the mutual information and the overall SNR is in very good agreement with our experimental measurements. The theoretical prediction of mutual information for the redundant measurement case requires knowledge of intrinsic–to–extrinsic noise ratio (IER). IER was estimated in two ways: by (i) quantifying the fluctuations in the later (quasistationary) portion of the response time series of our ERK data (fig. S21A) (ii) using data for repeated measurements of single-cell responses (*24*) (fig. S21B). The predicted mutual information based on redundant responses required IER values that are two to four orders of magnitude higher than experimentally estimated IER values (SM section 4.3.3) (fig S17). In contrast, the measured mutual information values were in good agreement with the theoretical prediction for a dynamic response based on a computational ERK model (SM section 4.1) (fig. S18). Overall, this analysis demonstrates that the substantial information gain from multivariate measurements is indeed the direct result of the dynamic nature of ERK response.

The robustness of biological systems is epitomized by their ability to function in the presence of a large variability in cellular states (*25*, *26*). Signaling dynamics allow biochemical networks to mitigate variability in the cellular state and thereby maximize the information transmission capacity of signaling networks. Although the theory and observations presented here focus on the information transmission capacity of the dynamics of a single signaling molecule, the extension of our analysis to the case of multiple signaling molecules responding to one ligand is straightforward. Not all of the information contained in the dynamical responses may actually

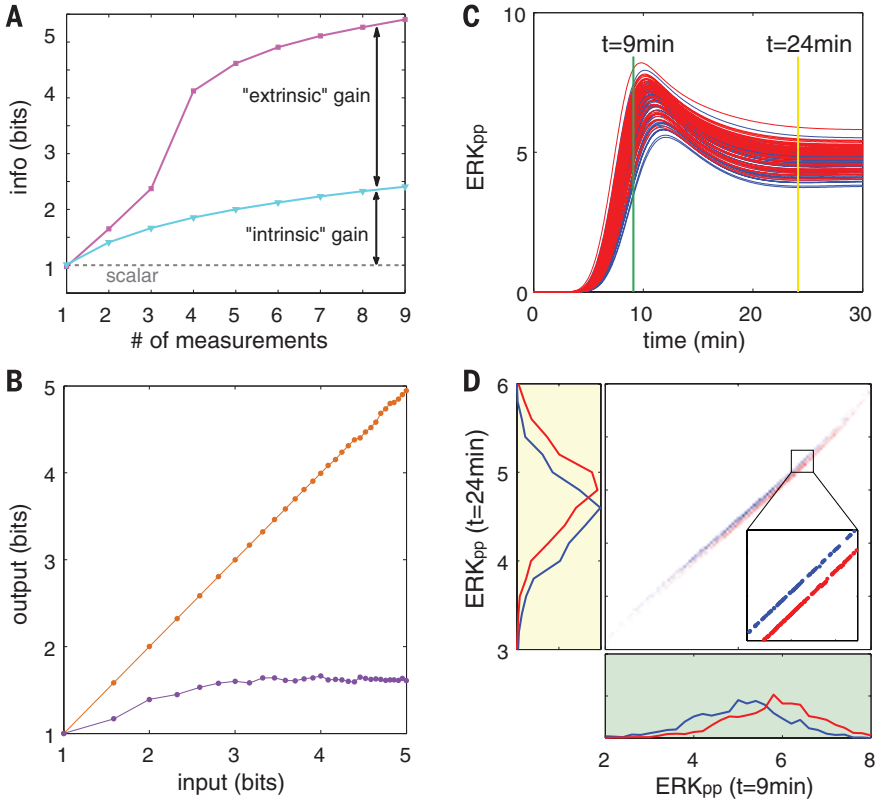


Fig. 3. Theoretical decomposition of information loss caused by intrinsic and extrinsic noise. (**A**) Graphical representation of the analytical expression for the gain in mutual information from overcoming intrinsic (cyan) and extrinsic (magenta) noise sources obtained from random linear Gaussian inputs and outputs with three parameters (*19*). (**B**) Information transmission capacity of dynamic (orange) and static (maximal response, purple) responses calculated using simulated trajectories from the computational model of ERK (*22*) with only the extrinsic noise contributing to cell response variability. (**C**) Example of ERK trajectory variability for two different inputs levels (red and blue). Variability was generated using a uniform distribution of a single parameter, MEK values, that was varied by ±20%. (**D**). Two-dimensional histogram (center) and marginal distributions (left and bottom) for the two input levels (shown in red and blue) at two time points (t = 9 and 24 min) from the trajectories in (C). Because only a single parameter was varied, the responses vary on a 1D curve. As a result, although the univariate marginal distributions show substantial response overlap, the 2D distribution shows completely separated response levels (inset).

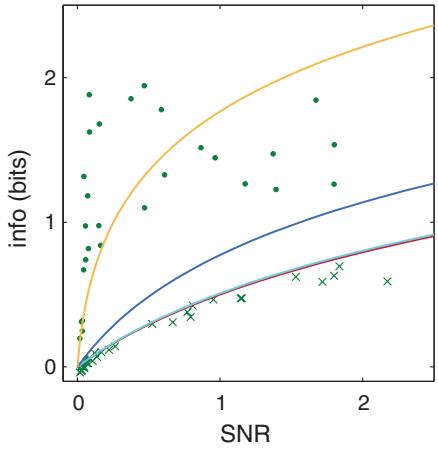


Fig. 4. Measured information gain is a result of ERK dynamics' ability to mitigate extrinsic noise. Experimental measurement of the mutual information between ERK response and EGF measured as a function of the response signal-to-noise ratio (SNR). Each marker represents calculations of SNR and mutual information from the dynamic (dot) and maximal scalar (cross) responses of cells from an eight-well dose-response experiment. Data shown are calculated based on 535,107 single-cell responses from 29 experiments with six doses of MEK inhibitor U0126 (tables S4 and S5). Lines represent theoretical predictions of the mutual information as a function of SNR for three types of responses: static scalar (red line), redundant measurements where the multivariate response has no dynamics (dark and light blue lines) calculated based on two independent estimates of IER (*19*) (fig. S21), and dynamic response (orange) that can mitigate both intrinsic and extrinsic noise.

be used by cells. Yet, because reliable information transmission is a fundamental function of cellular signaling networks, it is plausible that evolutionary pressures shaped the cellular machinery to maximize the reliable decoding of multivariate dynamic signals.

REFERENCES AND NOTES

1. M. D. Brennan, R. Cheong, A. Levchenko, *Science* **338**, 334–335 (2012).
2. A. Rhee, R. Cheong, A. Levchenko, *Phys. Biol.* **9**, 045011 (2012).
3. C. Waltermann, E. Klipp, *Biochim. Biophys. Acta* **1810**, 924–932 (2011).
4. F. Tostevin, P. R. ten Wolde, *Phys. Rev. Lett.* **102**, 218101 (2009).
5. W. H. de Ronde, F. Tostevin, P. R. ten Wolde, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **82**, 031914 (2010).
6. W. H. de Ronde, F. Tostevin, P. R. Ten Wolde, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **86**, 021913 (2012).
7. G. Tkacik, C. G. Callan Jr., W. Bialek, *Proc. Natl. Acad. Sci. U.S.A.* **105**, 12265–12270 (2008).
8. A. M. Walczak, G. Tkačik, W. Bialek, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **81**, 041905 (2010).
9. P. Mehta, S. Goyal, T. Long, B. L. Bassler, N. S. Wingreen, *Mol. Syst. Biol.* **5**, 325 (2009).
10. R. C. Yu *et al.*, *Nature* **456**, 755–761 (2008).
11. M. Behar, A. Hoffmann, *Biophys. J.* **105**, 231–241 (2013).
12. R. Cheong, A. Rhee, C. J. Wang, I. Nemenman, A. Levchenko, *Science* **334**, 354–358 (2011).
13. S. Uda *et al.*, *Science* **341**, 558–561 (2013).
14. M. Voliotis, R. M. Perrett, C. McWilliams, C. A. McArdle, C. G. Bowsher, *Proc. Natl. Acad. Sci. U.S.A.* **111**, E326–E333 (2014).
15. N. Hao, B. A. Budnik, J. Gunawardena, E. K. O'Shea, *Science* **339**, 460–464 (2013).
16. A. Hoffmann, A. Levchenko, M. L. Scott, D. Baltimore, *Science* **298**, 1241–1245 (2002).
17. S. D. M. Santos, P. J. Verveer, P. I. H. Bastiaens, *Nat. Cell Biol.* **9**, 324–330 (2007).
18. J. E. Purvis, G. Lahav, *Cell* **152**, 945–956 (2013).
19. T. M. Cover, J. A. Thomas, *Elements of Information Theory 2nd* (Wiley-Interscience, ed. 2, Hoboken, NJ 2006).
20. R. E. C. Lee, S. R. Walker, K. Savery, D. A. Frank, S. Gaudet, *Mol. Cell* **53**, 867–879 (2014).
21. C. Cohen-Saidon, A. A. Cohen, A. Sigal, Y. Liron, U. Alon, *Mol. Cell* **36**, 885–893 (2009).
22. O. E. Sturm *et al.*, *Sci. Signal.* **3**, ra90 (2010).
23. T. Schreiber, *Phys. Rep.* **308**, 1–64 (1999).
24. J. E. Toettcher, O. D. Weiner, W. A. Lim, *Cell* **155**, 1422–1434 (2013).
25. G. von Dassow, E. Meir, E. M. Munro, G. M. Odell, *Nature* **406**, 188–192 (2000).
26. N. Barkai, S. Leibler, *Nature* **387**, 913–917 (1997).

ACKNOWLEDGMENTS

We are indebted to T. Kemp (Univ. of California–San Diego) for discussion of the random Gram matrix theory. Supported by the NIH grants P50-GM085764 (R.W., A.H., L.T.), R01-GM089976 (L.T), and R01-GM071573 (A.H).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/346/6215/1370/suppl/DC1
Supplementary Text
Figs. S1 to S25
Tables S1 to S6
References (*27*–*38*)

17 April 2014; accepted 12 November 2014
10.1126/science.1254933

ONCOGENE REGULATION

An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element

Marc R. Mansour,[1,2] Brian J. Abraham,[3]* Lars Anders,[3]* Alla Berezovskaya,[1] Alejandro Gutierrez,[1,4] Adam D. Durbin,[1] Julia Etchin,[1] Lee Lawton,[3] Stephen E. Sallan,[1,4] Lewis B. Silverman,[1,4] Mignon L. Loh,[5] Stephen P. Hunger,[6] Takaomi Sanda,[7] Richard A. Young,[3,8]† A. Thomas Look[1,4]†

In certain human cancers, the expression of critical oncogenes is driven from large regulatory elements, called super-enhancers, that recruit much of the cell's transcriptional apparatus and are defined by extensive acetylation of histone H3 lysine 27 (H3K27ac). In a subset of T-cell acute lymphoblastic leukemia (T-ALL) cases, we found that heterozygous somatic mutations are acquired that introduce binding motifs for the MYB transcription factor in a precise noncoding site, which creates a super-enhancer upstream of the *TAL1* oncogene. MYB binds to this new site and recruits its H3K27 acetylase–binding partner CBP, as well as core components of a major leukemogenic transcriptional complex that contains RUNX1, GATA-3, and TAL1 itself. Additionally, most endogenous super-enhancers found in T-ALL cells are occupied by MYB and CBP, which suggests a general role for MYB in super-enhancer initiation. Thus, this study identifies a genetic mechanism responsible for the generation of oncogenic super-enhancers in malignant cells.

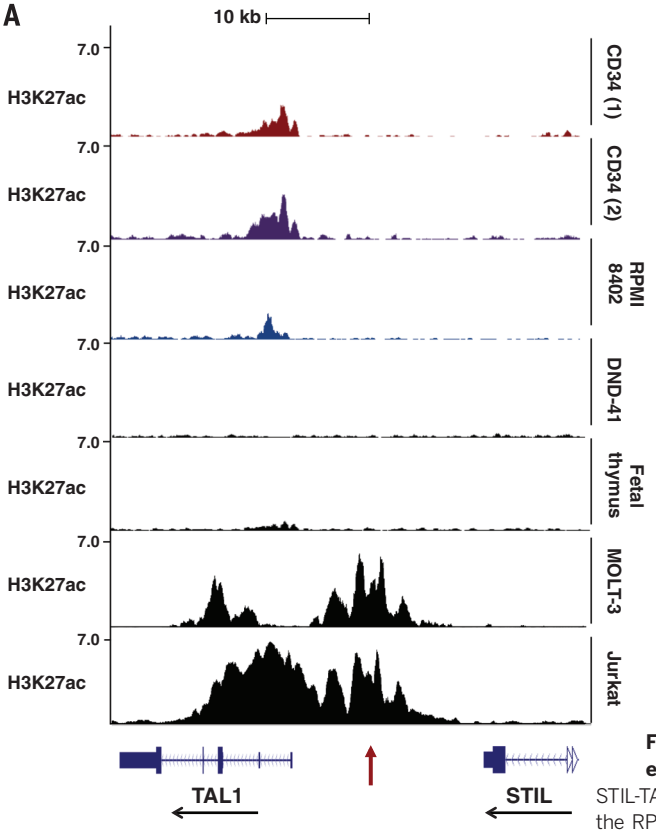
In cancer cells, monoallelic expression of oncogenes can occur through a variety of mechanisms, including chromosomal translocation, alterations in promoter methylation, parental imprinting, and intrachromosomal deletion (*1*–*3*). A quintessential example is TAL1^{d}, an ~80–kilobase (kb) deletion on chromosome 1p33 that is found in 25% of cases of human T cell acute lymphoblastic leukemia (T-ALL). The deletion results in overexpression of *TAL1*, an oncogene coding for a basic helix-loop-helix transcription factor, by mediating fusion of *TAL1* coding sequences to the regulatory elements of the ubiquitously expressed gene "*SCL*-interrupting locus" (*STIL*) (*4*–*6*). However, we previously reported that a substantial proportion of T-ALLs, including the Jurkat T-ALL cell line, have monoallelic overexpression of *TAL1* but lack either the TAL1^{d} abnormality or a chromosomal translocation of the *TAL1* locus (*7*, *8*).

We hypothesized that cis-acting genomic lesions affecting *TAL1* regulatory sequences might account for monoallelic *TAL1* activation. Chromatin immunoprecipitation (ChIP)–sequencing (ChIP-seq) analysis of Jurkat cells revealed aberrant histone H3 lysine 27 acetylation (H3K27ac), a mark of active transcription, starting upstream of the *TAL1* transcriptional start site and extending across the first exons (Fig. 1A) (*9*, *10*). Regions with such rich and broad H3K27ac marks have been termed super-enhancers (also stretch enhancers or locus control regions) and are commonly found at genes that determine cell identity in embryonic stem (ES) cells and in tumor cells at oncogenes critical for the malignant cell state (*11*–*17*). The super-enhancer encompassing *TAL1* in Jurkat cells was aberrant, in that it was not present in fetal thymocytes, normal CD34+ hematopoietic stem and progenitor cells (HSPCs), or in other T-ALL cell lines, such as TAL1^{d}-positive RPMI-8402 cells and DND-41 T-ALL cells that lack *TAL1* expression (Fig. 1A) (*9*). Of note, chromatin conformation capture experiments recently performed in Jurkat cells identified a looping interaction involving an enhancer site 8 kb upstream of the transcription start site (TSS), which coincides with the locations of both the aberrant super-enhancer and the positive autoregulatory binding sites for members of the TAL1 complex in this cell line (Fig. 1A, red arrow) (*9*, *18*).

Sequencing of the genomic DNA region encompassing this site identified a heterozygous 12–base pair (bp) insertion (GTTAGGAAACGG) that aligned precisely with the TAL1, GATA3, RUNX1, and HEB ChIP-seq peaks (Fig. 1B). Among eight additional TAL1-positive T-ALL cell lines, MOLT-3 cells also harbored an abnormal heterozygous 2-bp insertion (GT) at the same site (Fig. 1B), whereas none of 10

[1]Department of Pediatric Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA 02215, USA. [2]Department of Haematology, UCL Cancer Institute, University College London, London WC1E 6BT, UK. [3]Whitehead Institute for Biomedical Research, Cambridge, MA 02142, USA. [4]Division of Pediatric Hematology-Oncology, Boston Children's Hospital, MA 02115, USA. [5]Department of Pediatrics, Benioff Children's Hospital, University of California San Francisco, CA 94143, USA. [6]Pediatric Hematology/Oncology/BMT, University of Colorado School of Medicine and Children's Hospital Colorado, Aurora, CO 80045, USA. [7]Cancer Science Institute of Singapore, National University of Singapore, and Department of Medicine, Yong Loo Lin School of Medicine, 117599, Singapore. [8]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142, USA.
*These authors contributed equally and are listed in alphabetical order. †Corresponding author. E-mail: thomas_look@dfci.harvard.edu (A.T.L.); young@wi.mit.edu (R.A.Y.)

TAL1-negative cell lines had a detectable genomic abnormality in this region (table S1). Among 146 unselected pediatric primary T-ALL samples collected at diagnosis, eight patients (5.5%) had heterozygous indels 2 to 18 bp in length that overlapped at the same clearly defined hotspot (indels at this site are referred to here as "mutation of the TAL1 enhancer," MuTE) (Fig. 1B). We estimate that MuTE abnormalities account



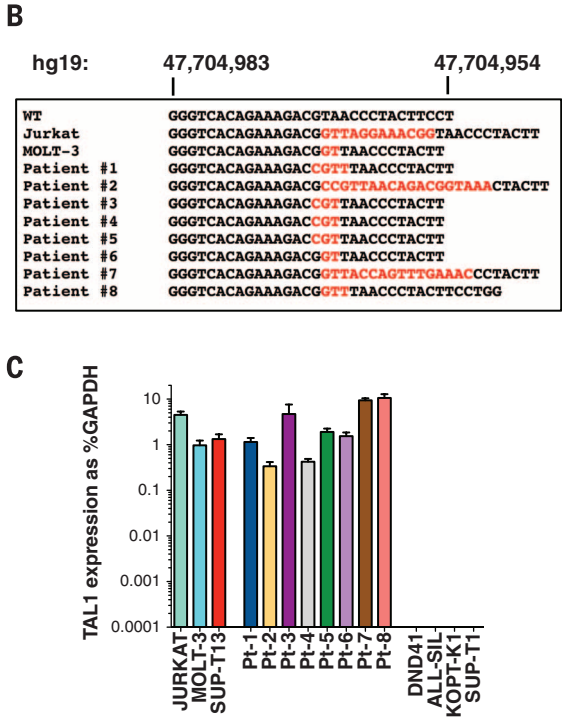
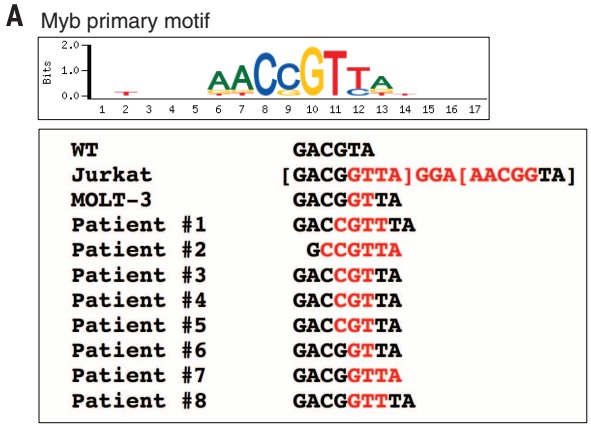


Fig. 1. Mutations at an intergenic site are associated with the *TAL1* super-enhancer in T-ALL cells. (**A**) Normalized ChIP-seq tracks for H3K27ac at the STIL-TAL1 locus in two human purified normal hematopoietic stem cell samples (CD34), the RPMI-8402 T-ALL cell line that overexpresses *TAL1* as a result of TAL1^{d} (RPMI-8402 cells), DND-41 T-ALL cells that do not express *TAL1*, human fetal thymic tissue, and MOLT-3 and Jurkat cells that have mutations at a noncoding site 7.5 kb from the *TAL1* transcriptional start site (red arrow). Black arrows indicate the direction of transcription. ChIP-seq read densities (*y* axis) were normalized to reads per million reads sequenced in each sample. (**B**) Sequence alignments of the −7.5 kb site showing wild-type (WT) sequences in black and inserted sequences in red for Jurkat and MOLT-3 T-ALL cell lines and eight pediatric T-ALL patients. hg19, human genome build 19. (**C**) *TAL1* mRNA expression as determined by quantitative polymerase chain reaction (PCR) and expressed as percentage of glyceraldehyde-3-phosphate dehydrogenase (GAPDH). Error bars are ± SEM from two independent experiments performed in triplicate.



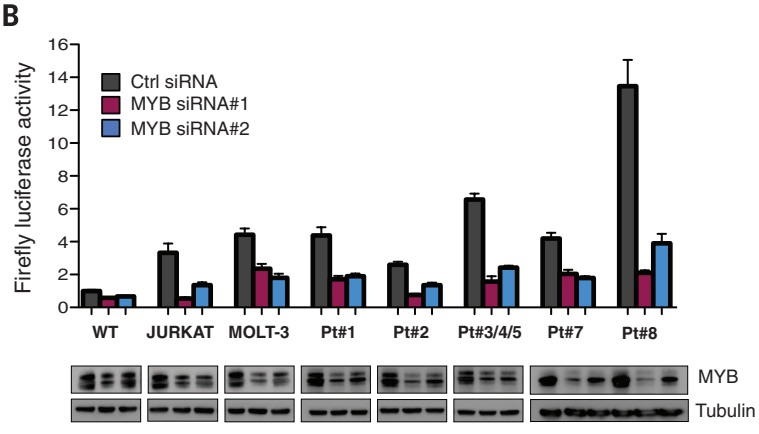


Fig. 2. Mutations of the *TAL1* enhancer activate through recruitment of MYB. (**A**) All *TAL1* enhancer mutations introduce de novo MYB binding sites as determined by UniPROBE (*19*). The MYB primary binding motif is shown above the mutation-derived MYB motifs, with inserted nucleotides shown in red. (**B**) A 400-bp fragment of the −7.5 kb *TAL1* enhancer containing the wild-type sequence or each of the mutant alleles was cloned upstream of luciferase and a minimal promoter. Constructs were delivered by the momentary creation of small pores in cell membranes by applying an electrical pulse into Jurkat cells (electroporation), together with either control siRNA or two independent siRNAs targeting MYB. Firefly luciferase activity was measured at 24 hours, normalized to renilla luciferase to control for cell number and transfection efficiency, and expressed as a ratio relative to activity of the wild-type enhancer construct. Error bars are ± SEM from two independent experiments performed in triplicate. Corresponding immunoblots for MYB and tubulin are shown below.

for about half of the cases with unexplained monoallelic overexpression of *TAL1* (*7*). Sequencing of DNA from remission bone marrow samples available from two mutation-positive patients showed wild-type sequences at this site, which indicated that the mutations were somatically acquired in the tumor cells. All eight MuTE-positive T-ALLs markedly overexpressed *TAL1* mRNA at levels comparable to those of the TAL1^{d}-positive SUP-T13, and MuTE-positive Jurkat and MOLT-3 cells (Fig. 1C). Furthermore, MOLT-3 cells that share the same 2-bp insertion seen in patient 6 also had a super-enhancer at the *TAL1* locus (Fig. 1A).

To test our hypothesis that the aberrantly formed super-enhancer drives monoallelic *TAL1* expression, we analyzed the MuTE-positive samples for single-nucleotide polymorphisms in the 3′ untranslated region (3′ UTR) of the *TAL1* gene from genomic DNA. Jurkat cells and five of the patient samples were informative, and in each case, only one allele was detectable in the RNA sample, which indicated monoallelic expression of *TAL1* (fig. S1). We then used the UniPROBE database to analyze whether the mutant sequences introduced new transcription factor binding sites (*19*). Note that all of the indels at this hotspot introduced de novo binding motifs for the MYB transcription factor, and two consecutive MYB binding motifs were generated by the 12-bp insertion in Jurkat cells (Fig. 2A and table S2).

To ascertain whether these mutations can activate gene expression, we cloned a 400-bp fragment containing either the wild-type allele or each of the *TAL1*-enhancer mutant alleles upstream of luciferase and tested the enhancer activity of this fragment in reporter assays. When we expressed these constructs in Jurkat cells, fragments containing each of the seven different mutations robustly increased reporter activity 3- to 14-fold as much as the wild-type fragment (Fig. 2B and fig. S2). Moreover, the activity of each of the mutant reporters was markedly reduced after MYB knockdown, which indicated that the enhancer activity imparted by the mutations was indeed mediated by MYB (Fig. 2B). When we performed these experiments in human embryonic kidney 293T (HEK-293T) cells, the mutant reporters had no increased activity above that of the wild-type reporter (fig. S2), which suggested that transcription factors expressed in T-ALL, such as members of the TAL1 complex, are involved in activation of the mutant enhancer. We conclude that, in T-ALL primary samples and cell lines, indel mutations that introduce MYB binding sites at a hotspot 7.5 kb upstream from the TSS of *TAL1* generate a super-enhancer that drives monoallelic overexpression of this oncogene.

Our recent ChIP-seq studies of the TAL1 complex in T-ALL cells identified binding of transcription factors in the core TAL1 complex (TAL1, GATA3, RUNX1, E2A, and HEB) at the MYB enhancer, whereas knockdown of MYB generated a gene expression signature closely related to TAL1 knockdown (*9*, *10*). We were technically unable to analyze MYB binding by ChIP-seq in our previous study, so at that time, we interpreted these results to indicate that MYB is a critical downstream hub of the TAL1 complex (*10*). However, in the current study we used newly available MYB-specific antibodies to generate high-resolution maps of genome-wide MYB binding in Jurkat cells. Analysis of the TAL1 enhancer indel mutation site in Jurkat cells showed precise alignment of MYB binding and binding of each member of the TAL1 complex (Fig. 3A). There was also an abundance of RNA polymerase II (Pol II) and Mediator (MED1), stretching over more than 20 kb, which indicated a large super-enhancer (table S3) (*15*). Notably, this site is also bound by MYB and TAL1 in MuTE-positive MOLT-3 cells (Fig. 3A) but not in RPMI-8402 and CCRF-CEM cells or a primary T-ALL sample, each of which overexpresses TAL1 driven by TAL1^{d} (fig. S3). Nor were we able to detect binding of TAL1 at this site in HSPCs (fig. S3), which indicated that a small insertion creating a de novo MYB binding motif at this location is required for MYB binding and subsequent binding by other members of the TAL1 complex. Accordingly, knockdown of MYB resulted in depletion of *TAL1* expression in both Jurkat and MOLT-3 cells (fig. S4). Thus, MYB binding to the MuTE hotspot in a subset of *TAL1*-overexpressing T-ALLs results in the accumulation of an abundance of H3K27ac marks and aberrantly nucleates binding by other members of the TAL1 complex, leading to aberrant up-regulation of *TAL1* gene expression.

We next asked why the mutations we had identified in primary patient T-ALLs were clustered in a defined genomic location. A search for predicted transcription factor binding sites near the MuTE site identified the preferred binding sequences for RUNX1, GATA3, and ETS1, as well as E-box motifs characteristic of binding by TAL1/E-protein heterodimers (fig. S5). The absence of predicted MYB binding sites in the wild-type sequence suggests that the MuTE is critical for MYB binding and supports our hypothesis that MYB binding to its de novo motif is crucial to binding by members of the TAL1 complex at this hotspot. To explore this concept further, we extracted the raw ChIP-seq reads and determined the allelic frequency of

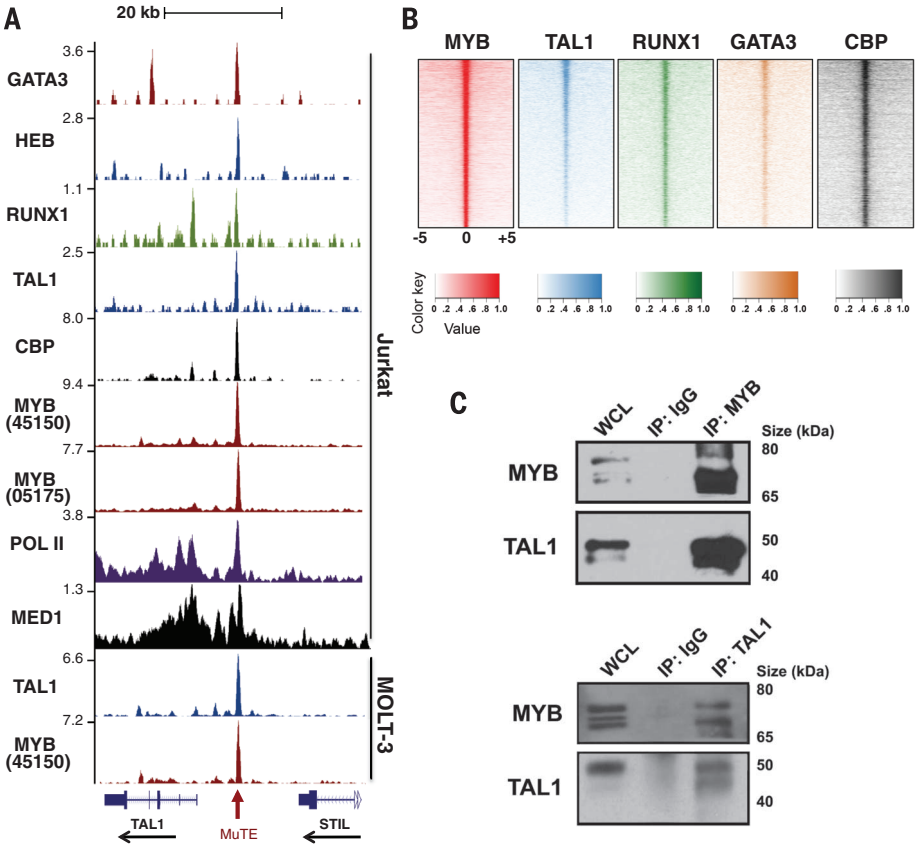


Fig. 3. MYB binds the mutant *TAL1* enhancer site and is a member of the TAL1 complex. (**A**) ChIP-seq tracks at the STIL-TAL1 locus from Jurkat and MOLT-3 T-ALL cells for GATA3, HEB, RUNX1, TAL1, CBP, MYB (ab45150 antibody), MYB (05-175 antibody), RNA polymerase II (Pol II), and Mediator 1 (MED1). The mutation site at −7.5 kb is depicted with a red arrow. ChIP-seq read densities (*y* axis) were normalized to reads per million reads sequenced in each sample. (**B**) Heat maps showing genome-wide cooccupancy of MYB binding sites (±5 kb) with those from TAL1, RUNX1, GATA3, and CBP sites as determined by ChIP-seq. For each region (*y* axis), the read density centered at 0 indicates overlapping bound regions. (**C**) Co-IP and reciprocal Co-IP experiments performed from Jurkat lysates for MYB and TAL1. WCL, whole-cell lysate; IgG, isotype control immunoglobulin G antibody.

mutant to wild-type reads of bound DNA fragments at the mutation site. Note that, in MOLT-3 cells, both MYB and TAL1 bound predominantly to the mutant allele, with 67 of the 68 reads, and 37 of 38 reads, revealing the mutant sequence in the bound DNA. Likewise, in Jurkat cells, 404 of 419 reads, and 12 of 14 reads, contained the mutant sequence from MYB and TAL1 ChIP, which indicated that these transcription factors predominantly bind monoallelically to the mutant allele. Thus, our data indicate that indels producing MYB binding sites occur at a defined genomic location, probably because they must be in proximity to binding sites for other members of the TAL1 complex.

Given the well-established direct interaction between MYB and its potent transcriptional coactivator, CREB-binding protein (CBP) (*20*), we also analyzed ChIP-seq tracks for CBP in Jurkat cells and saw that CBP was also present at the *TAL1* enhancer indel mutation site (Fig. 3A). Because CBP promotes H3K27 acetylation and antagonizes Polycomb silencing (*21*), it seems likely that the initial event in the aberrant super-enhancer formation at this site is the recruitment of CBP by MYB, resulting in abundant H3K27 acetylation, which in turn opens the chromatin and permits binding by the other members of the TAL1 complex.

In a sense, the mutations forming MYB binding sites upstream of the TAL1 gene in T-ALLs represent an "experiment of nature" that reveals the capacity of MYB binding to aberrantly nucleate a large super-enhancer that drives high levels of expression of a gene critical for the leukemic cell state. Thus, we interrogated the normal binding sites of MYB and observed highly concordant binding of MYB with members of the TAL1 complex throughout the genome (Fig. 3B), in that 80% of TAL1 binding sites were also cooccupied by MYB. We were also able to demonstrate a strong association of TAL1 and MYB proteins biochemically in reciprocal coimmunoprecipitation experiments (Fig. 3C). Next, we focused on the positive interconnected autoregulatory loop that we had identified previously, whereby the core components of the TAL1 complex, TAL1, GATA3, and RUNX1, positively regulate their own enhancers (*9*). We found MYB bound with TAL1, GATA3, and RUNX1 at each of their respective enhancers, including one within the *MYB* gene itself (fig. S6). Notably, the TAL1 complex binding sites associated with all four of these genes also contained large super-enhancer domains, and all showed CBP-MYB cooccupancy. Furthermore, small interfering RNA (siRNA) knockdown of MYB was sufficient to deplete TAL1, GATA3, and RUNX1 (fig. S6). Thus, MYB is not only a core component of the TAL1 complex but also a key factor involved in initiating the autoregulatory positive-feedback circuitry (fig. S6).

To demonstrate definitively that MuTEs are responsible for TAL1 overexpression in a subset of T-ALL patients, we used clustered regularly interspaced short palindromic repeats (CRISPR)–associated (CRISPR-Cas9) technology to disrupt the MuTE site in Jurkat cells. Initially, we had difficulty expanding single-cell clones with deletion of the MuTE site, which suggested that deletion of the mutated enhancer site was diminishing TAL1 levels to a degree that impaired cell survival. Thus, we engineered Jurkat cells to express *TAL1* cDNA lacking the 3′ UTR from a retroviral vector and performed all of our CRISPR-Cas9 experiments in these cells.

To directly target the enhancer mutation site in Jurkat cells, we first generated clones with a deletion of ~180 bp by targeting two guide RNAs to sites flanking either side of the enhancer mutation. Clones expanded from single cells harbored genomic deletions of 177 to 193 bp that involved the wild-type allele or mutant allele or both alleles (Fig. 4A and fig. S7). Deletion of the wild-type allele had no effect on endogenous *TAL1* mRNA levels, but deletion of the mutant allele completely abrogated endogenous *TAL1* expression, which indicated that the enhancer mutation

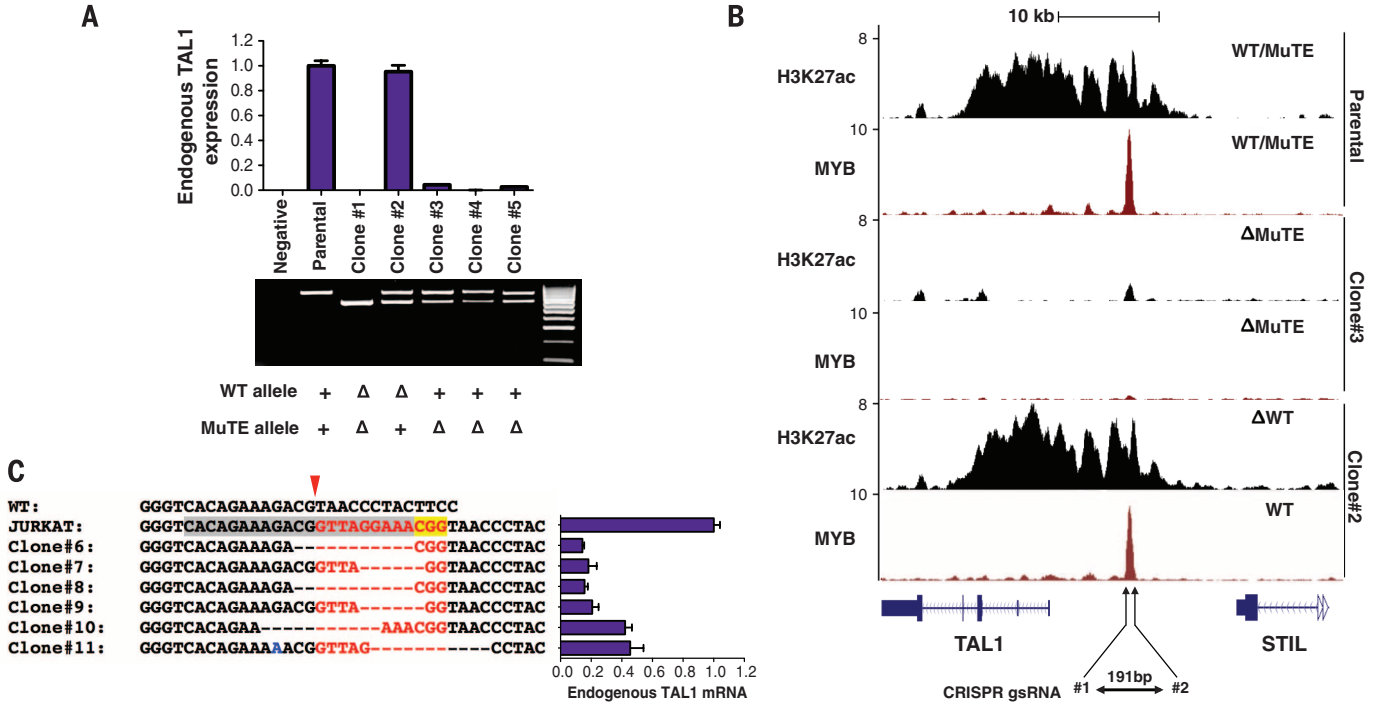


Fig. 4. Targeted deletion of the *TAL1* enhancer mutation collapses the *TAL1* super-enhancer. (**A**) Targeted deletion of 177 to 193 bp of the mutant, but not wild-type, allele in Jurkat cells abrogates expression of endogenous TAL1, as determined by quantitative reverse transcription PCR (qRT-PCR). Data are means ± SD of two independent experiments performed in triplicate. Agarose gel of products from PCR amplification across the MuTE site for CRISPR-Cas9 Jurkat clones. All clones, including parental cells, express MSCV-TAL1. Hyperladder IV on right. Genotype for each clone is shown below: +, allele present; Δ, deleted allele. (**B**) ChIP-seq tracks for H3K27ac and MYB at the STIL-TAL1 locus from selected CRISPR-Cas9 clones. ChIP-seq read densities (*y* axis) were normalized to reads per million reads sequenced in each sample. (**C**) Sequence alignments of Jurkat clones targeted by CRISPR-Cas9 germline small RNA (gsRNA) number 3 [target sequence is highlighted in gray, protospacer adjacent motif (PAM) sequence in yellow], which targets the 12-bp insertion in Jurkat cells (red font) but not the wild-type allele. Endogenous TAL1 expression as determined by qRT-PCR for respective clones is shown. Data are means ± SD of two independent experiments performed in triplicate.

is responsible for *TAL1* overexpression in these cells. Furthermore, ChIP-seq for H3K27ac showed complete collapse of the super-enhancer at the *TAL1* locus when the deletion affected the allele with the enhancer mutation but was not affected when the deletion involved only the wild-type allele (Fig. 4B and figs. S8 and S9).

We also targeted a single guide RNA to specific sequences that form part of the 12-bp insertion in Jurkat cells, which permitted us to propagate single-cell clones with a spectrum of repair-induced indel mutations directly at the insertion site (Fig. 4C). In clones with deletion of 6 bp of the 12-bp enhancer insertion, encompassing one of the two inserted MYB binding sites, endogenous TAL1 expression levels decreased by ~60%, whereas clones with more extensive deletions had endogenous TAL1 expression levels decreased by ~85% (Fig 4C). Thus, the MuTE is clearly responsible for TAL1 overexpression in Jurkat cells.

Our ChIP-seq results also show that MYB and CBP were bound together at 727 of the 818 (89%) super-enhancer regions that are present in Jurkat cells. When we performed short hairpin RNA–mediated knockdown for MYB, 221 of 818 (27%) super-enhancer–associated genes decreased significantly in expression (*9*, *17*), which suggested that MYB has an active role in regulating their transcription. These results are consistent with the interpretation that MYB-CBP binding and the subsequent formation of abundant H3K27 acetylation marks may be broadly involved in the formation of super-enhancers in T-ALL. Thus, the role that we have shown for MYB binding in super-enhancer formation in a subset of T-ALLs with strategically placed somatic indel mutations in all likelihood provides insight into the general question of how super-enhancers are formed at the site of genes critical for the establishment of the T-ALL cell state. MYB is known to function as a master regulator of early and adult hematopoiesis and to undergo transcriptional down-regulation after lineage commitment and differentiation (*22*). An interesting area for future study will be to determine whether MYB acts in concert with CBP to regulate super-enhancer formation at genes critical for defining cell identity during normal hematopoietic cell differentiation (*14*, *16*, *23*, *24*).

Our findings show that somatic mutation of noncoding intergenic elements can lead to binding of master transcription factors, such as MYB, which in turn aberrantly initiate super-enhancers that mediate overexpression of oncogenes. This raises the possibility that acquisition of such enhancer mutations may constitute a general mechanism of carcinogenesis used in other types of human cancers. Mechanisms of aberrant super-enhancer formation in malignancy have broad implications not only for molecular pathogenesis but also for clinical management. Drugs that target key components of the transcriptional machinery, such as BRD4 and CDK7 (*12*, *13*, *17*), have recently been shown to preferentially target tumor-specific super-enhancers, which provides a novel strategy to capitalize on these abnormalities for improved cancer therapy.

REFERENCES AND NOTES

1. E. J. Walker *et al.*, *Cancer Res.* **72**, 636–644 (2012).
2. A. Gimelbrant, J. N. Hutchinson, B. R. Thompson, A. Chess, *Science* **318**, 1136–1140 (2007).
3. R. L. Jirtle, *Exp. Cell Res.* **248**, 18–24 (1999).
4. T. M. Breit *et al.*, *J. Exp. Med.* **177**, 965–977 (1993).
5. L. Brown *et al.*, *EMBO J.* **9**, 3343–3351 (1990).
6. P. D. Aplan *et al.*, *Science* **250**, 1426–1429 (1990).
7. A. A. Ferrando *et al.*, *Blood* **103**, 1909–1911 (2004).
8. K. Leroy-Viard, M. A. Vinit, N. Lecointe, D. Mathieu-Mahul, P. H. Roméo, *Blood* **84**, 3819–3827 (1994).
9. T. Sanda *et al.*, *Cancer Cell* **22**, 209–221 (2012).
10. M. R. Mansour *et al.*, *J. Exp. Med.* **210**, 1545–1557 (2013).
11. R. Sierbæk *et al.*, *Cell Rep.* **7**, 1443–1455 (2014).
12. J. Lovén *et al.*, *Cell* **153**, 320–334 (2013).
13. P. Filippakopoulos *et al.*, *Nature* **468**, 1067–1073 (2010).
14. S. C. Parker *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 17921–17926 (2013).
15. D. Hnisz *et al.*, *Cell* **155**, 934–947 (2013).
16. W. A. Whyte *et al.*, *Cell* **153**, 307–319 (2013).
17. N. Kwiatkowski *et al.*, *Nature* **511**, 616–620 (2014).
18. Y. Zhou *et al.*, *Blood* **122**, 4199–4209 (2013).
19. D. E. Newburger, M. L. Bulyk, *Nucleic Acids Res.* **37** (Database), D77–D82 (2009).
20. P. Dai *et al.*, *Genes Dev.* **10**, 528–540 (1996).
21. F. Tie *et al.*, *Development* **141**, 1129–1139 (2014).
22. C. Schulz *et al.*, *Science* **336**, 86–90 (2012).
23. A. Rada-Iglesias *et al.*, *Nature* **470**, 279–283 (2011).
24. M. P. Creyghton *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 21931–21936 (2010).

ACKNOWLEDGMENTS

We thank J. Gilbert for helpful editorial comments on the manuscript, J. Reddy for sharing data, and F.Alt and F. Meng for helpful advice on the design of the CRISPR experiments. We would like to dedicate this paper to the memory of Michael Fayngersh. We gratefully acknowledge the children with T-ALL and their families for the samples analyzed in these studies. M.R.M. was supported by the Claudia Adams Barr Innovative Basic Science Research Program, the Kay Kendall Leukaemia Trust, and a Bennett Fellowship from Leukaemia and Lymphoma Research, UK. A.G. is a Research Fellow of the Gabrielle's Angel Foundation for Cancer Research, a Clinical Investigator of the Damon Runyon Cancer Research Foundation, and is supported by grants National Cancer Institute, NIH, CA167124, Department of Defense, CA120215 and an award from the William Lawrence and Blanche Hughes Foundation. T.S. is supported by a grant from the National Research Foundation (NRF), Prime Minister's Office, Singapore under its NRF Fellowship Program (award no. NRF-NRFF2013-02). This work was funded by NIH grants 1R01CA176746-01 and 5P01CA109901-08 (A.T.L.), and 5P01CA68484 (S.E.S. and A.T.L.). Children's Oncology Group cell banking and sample distribution were supported by grants CA98543, CA114766, CA98413, CA30969, and CA29139 from the NIH. S.P.H. is the Ergen Family Chair in Pediatric Cancer. R.A.Y. is a founder and member of the Board of Directors of Syros Pharmaceuticals, a company developing therapies that target gene regulatory elements including super-enhancers.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/346/6215/1373/suppl/DC1
Materials and Methods
Figs. S1 to S9
Tables S1 to S3
References (*25*–*44*)

22 July 2014; accepted 4 November 2014
Published online 13 November 2014;
10.1126/science.1259037

ELECTRON MICROSCOPY

Ultrastable gold substrates for electron cryomicroscopy

Christopher J. Russo and Lori A. Passmore*

Despite recent advances, the structures of many proteins cannot be determined by electron cryomicroscopy because the individual proteins move during irradiation. This blurs the images so that they cannot be aligned with each other to calculate a three-dimensional density. Much of this movement stems from instabilities in the carbon substrates used to support frozen samples in the microscope. Here we demonstrate a gold specimen support that nearly eliminates substrate motion during irradiation. This increases the subnanometer image contrast such that α helices of individual proteins are resolved. With this improvement, we determine the structure of apoferritin, a smooth octahedral shell of α-helical subunits that is particularly difficult to solve by electron microscopy. This advance in substrate design will enable the solution of currently intractable protein structures.

Recent developments in electron cryomicroscopy (cryo-EM) have allowed structure determination to near-atomic resolution for some macromolecular complexes (*1*–*3*). Still, many small and challenging structures cannot be determined by current cryo-EM methods to the resolutions required for accurate modeling of atom positions. This is because electron micrograph quality still falls short of the physical limits imposed by radiation damage to the macromolecules (*4*–*6*). Reduced image quality likely has two primary origins: inefficient detection of imaging electrons (*6*, *7*) and motion and blurring of the particles during irradiation (*6*, *8*–*11*). Recent developments in electron detectors have addressed the first constraint (*12*, *13*) and have enabled the development of motion correction algorithms to ameliorate the effects of the second (*14*–*16*). Direct electron detectors can acquire images using fractions of the electron dose previously required, thus allowing the measurement of single-molecule positions in time. This allows accurate tracking of large ensembles of particles and offers a way to determine the physical origins of radiation-induced particle movement (*14*–*17*).

Medical Research Council (MRC) Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, UK.
***Corresponding author. E-mail: passmore@mrc-lmb.cam.ac.uk**

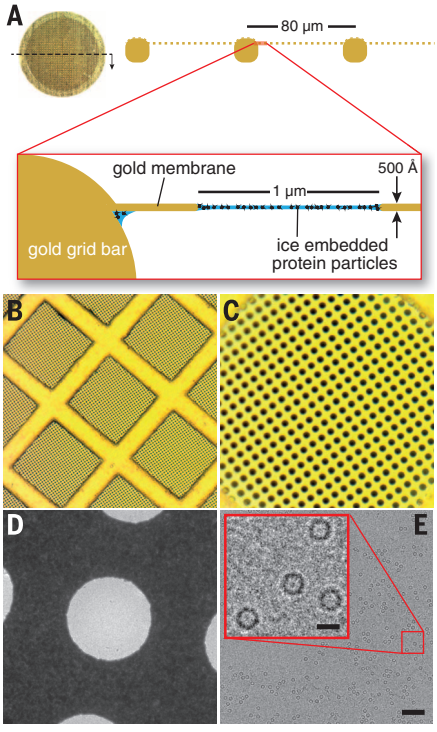


Fig. 1. Ultrastable substrate design. The ultrastable gold support comprises a 3-mm diameter disc of gold mesh (A) where a ≅500 Å thick layer of gold foil with a regular array of micrometer-sized holes is suspended across the square openings in the mesh (diagramed along the section indicated). After application of an aqueous protein sample and plunge freezing at ≅80 K, each hole contains a thin film of protein particles embedded in vitrified ice. (**A** to **C**) Optical micrographs of the gold grid at low, medium, and high magnification respectively; each hole is 1.2 μm in diameter and sets the scale for (B) to (D). (**D**) Transmission electron micrograph of an individual 1.2-μm hole containing vitrified ice. (**E**) A typical high-resolution electron micrograph of apoferritin suspended in ice on a gold grid. Scale bars for the micrograph and magnified inset are 1200 and 120 Å, respectively.

Conventional amorphous carbon (am-C) substrates undergo bending and deformation during irradiation (*9*, *11*), which includes movement both parallel and perpendicular to the substrate plane. Incorporation of titanium-silicon, doped silicon carbide, or graphene into substrate designs reduced radiation-induced specimen motion (*17*–*19*). Still, these designs did not stop substrate movement and are challenging to manufacture and use. Here we demonstrate a cryo-EM support that nearly eliminates the radiation-induced deformation of thin, ice-embedded specimens at cryogenic temperatures. This curtails the perpendicular and in-plane components of motion during imaging and thus improves image quality for all radiation-sensitive cryogenic specimens.

The support comprises a regular array of micrometer-sized holes in a ≅500 Å thick foil of gold (Fig. 1 and fig. S1; see also supplementary materials and methods). The foil is suspended across a mesh grid, also composed of gold, with square holes ≅80 μm wide. It differs from standard am-C supports only in the choice of materials and the thickness of the perforated foil. We chose gold because it is a highly conductive, nonoxidizing, radiation-hard material whose surface is chemically inert and biocompatible. Furthermore, making the foil and grid entirely out of the same metal ensures uniform electrical conductivity and prevents differential thermal contraction during cooling from 300 to 80 K, thus maintaining the geometry and tension of the support foil during use.

To characterize performance, we first measured the vertical motion of tilted gold substrates, without ice, under irradiation using standard cryogenic conditions (Fig. 2A and movie S1). Compared with commercial am-C supports with nearly identical geometry, there was a 60-fold reduction in substrate displacement (228 versus 3.8 Å) in a typical fluence used for high-resolution cryo-EM (16 e^-/Å^2). Next, we compared the vertical motion of the gold and am-C substrates when supporting a typical layer of thin ice used to image proteins (Fig. 2B). Adding the layer of ice stabilized both the am-C and gold supports, but there was still a 40-fold reduction in movement on gold compared with am-C (76 versus 1.9 Å).

To characterize the in-plane motion (parallel to the plane of the support) of proteins suspended in ice on gold substrates, we made test samples using 80*S* ribosomes and tracked them under standard imaging conditions. We analyzed a large ensemble of particle trajectories and compared these to previously published data on other substrates (Fig. 3A and fig. S2) (*17*). Particles generally exhibit two phases of motion during irradiation—a faster phase during the first 4e^-/Å^2 followed by a slower phase—and both are significantly reduced using gold (by ≅40% and ≅80%, respectively). This shows that much of the first phase and most of the second phase of particle motion are due to the support. On the gold substrate, the signal in the second part of the exposure approaches the physical limits imposed by damage and detector efficiency (fig. S3). Based on these experiments, we posit that reducing the vertical motion of the support by 50-fold reduces the total in-plane motion of the particles by twofold, where the coupling between the vertical and in-plane motion is due to bending of the irradiated region. We include a model of substrate deformation accounting for this in the supplement (fig. S4).

A reduction in motion directly improves the quality of the images. We quantify this using statistical analysis of tilt-pair images of ribosomes (*20*) collected on the different supports (fig. S5). These data demonstrate a 140% improvement in κ, which is a direct measure of the quality of the images. This corresponds to a 35% improvement in the accuracy of the angles assigned to the individual images during three-dimensional (3D) reconstruction (fig. S5C), which is crucial for solving macromolecular structures. Given a high-quality protein preparation, angular accuracy and the isotropic coverage of the information content in Fourier space are the primary factors that determine whether and how accurately a structure can be determined by cryo-EM.

To further assess the performance of gold supports, we determined the structure of apoferritin, a small protein (18 kD) that assembles into a

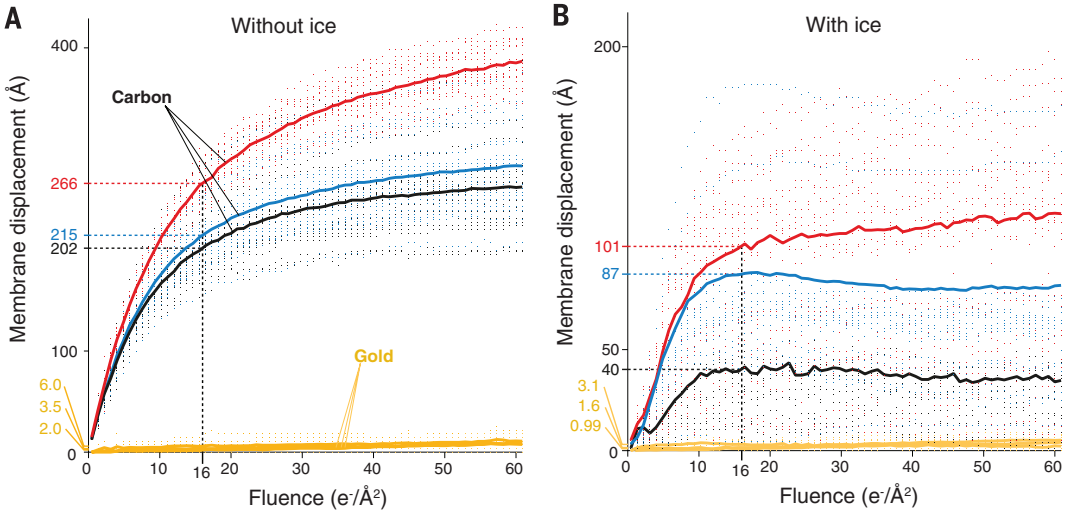


Fig. 2. Reduced motion of gold substrates under high-energy electron irradiation. (**A**) Measurements of the vertical motion of am-C and gold supports under typical cryo-EM illumination conditions (300 keV, 16 e^-/Å^2 per second, and 80 K). Each point is the magnitude of vertical displacement of the edge of a particular hole in the foil relative to its initial position before electron irradiation. Each solid line is the root mean square displacement for multiple holes in multiple squares of one grid. When a typical thin layer of vitreous ice is present (**B**), the vertical motion of both substrates is reduced about twofold, and the motion on am-C becomes more complicated in nature.

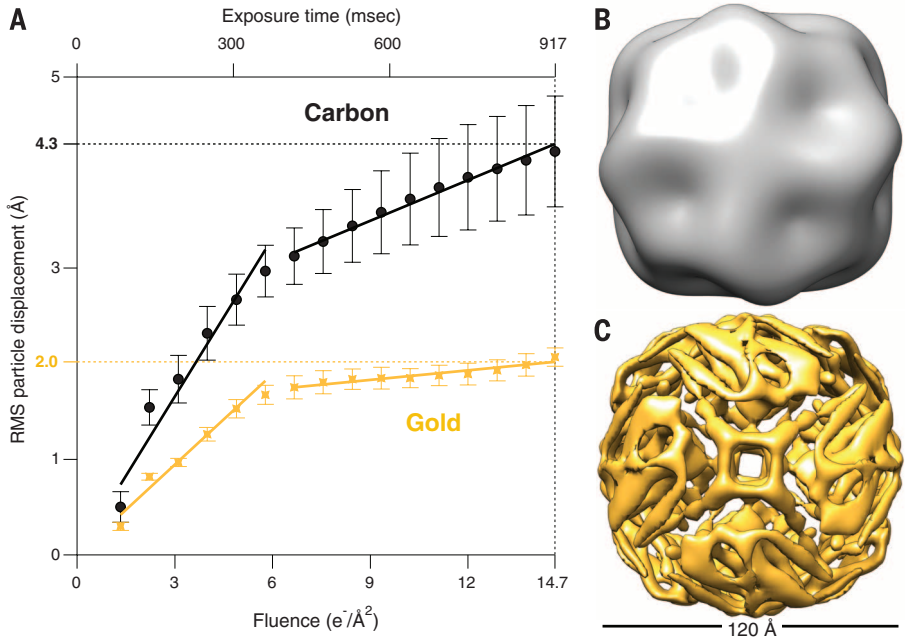


Fig. 3. Reduced particle motion and improved resolution on ultrastable grids. (**A**) 80*S* ribosomes were tracked during electron irradiation (using the same imaging conditions) for particles supported on three different am-C grids [black circles (*17*)] and three different gold grids (gold crosses). Each point is the in-plane ensemble average displacement of a particle from its initial position, and the error bars denote the SEM for the multiple grids tested. RMS, root mean square. Solid lines are linear fits to the two phases of motion, where the slopes are the average speeds of the particles. There was a 43% reduction in the speed of the first phase (8.8 to 5.0 Å/s) and a 77% reduction in the speed of the second phase (2.3 to 0.53 Å/s) on gold relative to am-C. Density maps of apoferritin processed identically from data collected using identical conditions for am-C (**B**) and gold (**C**) substrates. The map in (B) has a resolution of 25 Å (gold-standard Fourier shell correlation) and contains no discernible information beyond that present in the initial model. Compare to (C) with a resolution of 8.0 Å, which resolves the entire molecular chain.

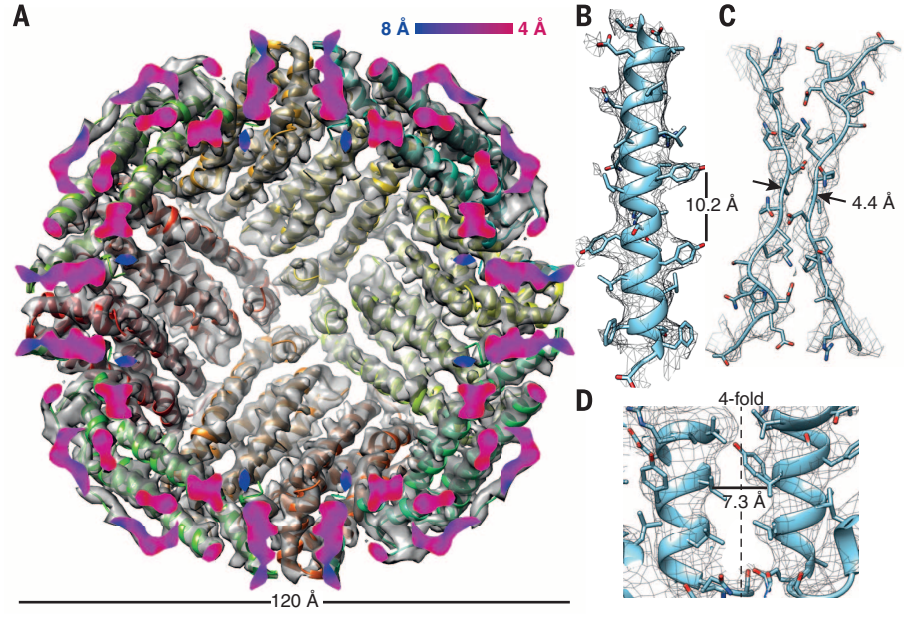


Fig. 4. The structure of apoferritin from 483 single-particle images. (**A**) The previously published x-ray crystal structure is fit to the cryo-EM density map, and the density is cut in half along the plane defined by two of the fourfold symmetry axes to reveal the interior of the complex. The cut surface shows the local resolution of the density using the color scale as indicated. (**B**) View of one of the α helices showing clear side-chain density. (**C**) View along the twofold symmetry axis showing clear separation between the β strands. (**D**) Section parallel to the fourfold pore axis showing the corrugated density of the residues in the interior of the pore.

smooth, spherical complex (440 kD). The ferritins are a class of iron-storage proteins that are conserved throughout evolution and whose characteristic structural motif is a bundle of four α helices (*21*, *22*). Apoferritin has remained intractable to structure determination by cryo-EM—even using the new generation of electron detectors—because the contrast in individual particle images was insufficient for the resolution of α helices, which is required to align the images with each other (*6*, *23*). We collected data on gold and am-C supports prepared identically (with identical geometry and imaging conditions, on the same day, and on the same microscope equipped with a back-thinned direct electron detector) and processed the resulting 4000 particles from each support type identically (Fig. 3, B and C). The images of apoferritin on am-C are still too blurred for correct alignment, resulting in a 3D density map that is no better than the initial model (resolution ≃25 Å); the reduced motion of the proteins on the gold grids improves the images enough that the orientations are correctly assigned, yielding a map that resolves the entire molecular chain (≃8 Å).

We used the gold support to collect a larger data set of ≃6000 particles. This yielded a reconstruction with a resolution of ≃7 Å (fig. S6). We suspected that conformational heterogeneity was limiting the resolution of the reconstruction. Using 3D classification (*24*), we were able to isolate a subset of 483 particles (11,592 asymmetric units) where a flexible loop on the exterior of the complex was positioned along the dimer interface, although we cannot exclude the possibility that the classification process may also have selected for particles that happened to move less during imaging. This improved the map (Fig. 4) and brought the resolution of the reconstruction to 4.7 Å (fig. S7).

The final map shows clearly identifiable density for the larger side chains, distinct separation of the β strands along the dimerization interface, and the molecular surface within the pore along the fourfold axis (Fig. 4, B to D). We found a region of extra density at the C terminus (fig. S8) corresponding to three amino acids in the protein sequence that were not modeled in the crystal structure (*25*). Refinement of an atomic model against the map also suggests an alternate conformation of the molecule relative to the crystal structure (fig. S7A).

The improved stability and image quality of gold substrates is not due to the mechanical strength of the suspended foil (fig. S9) but instead is likely due to preservation of the tension in the membrane during the cooling process, combined with the radiation hardness and high conductivity of the gold film at cryogenic temperatures. Substrates that reduce radiation-induced motion will improve the images from every microscope, not just those equipped with a direct electron detector or high-speed frame capture hardware and processing algorithms. We expect that the 50-fold reduction in the vertical motion of the gold substrates will also enable electron tomography at increased resolutions, as images of tilted specimens are more severely affected by the vertical motion of the substrate. The methods herein will allow one to solve the structures of many proteins previously refractory

to structural analysis, including other ferritins that could not be solved by x-ray crystallography.

During the first 4 $e^-/Å^2$ on gold substrates, 1 to 2 Å of in-plane motion remain. These first few electrons are critical, as they potentially contain the most high-resolution information (*26*, *27*). Future work will focus on substrate design and image acquisition conditions to reduce the initial motion even more. Along with further improvements in electron detector efficiency, this will bring cryo-EM to the physical limits imposed by the homogeneity of the protein preparation, the electron scattering cross sections of the biological specimen (*4*), and the counting statistics of detecting individual electrons. We anticipate that high-efficiency, stationary particle imaging will allow: (i) measurement and modeling of the progressive damage to the primary and secondary structure of the molecules, (ii) improved refinements using dose-fractionation that includes the use of tilt, and (iii) direct modeling and refinement of molecular structure from particle images. This will enable routine structure determination for many molecules and complexes that are currently too difficult to be practical.

REFERENCES AND NOTES

1. A. Amunts *et al.*, *Science* **343**, 1485–1489 (2014).
2. M. Liao, E. Cao, D. Julius, Y. Cheng, *Nature* **504**, 107–112 (2013).
3. M. Allegretti, D. J. Mills, G. McMullan, W. Kühlbrandt, J. Vonck, *eLife* **3**, e01963–e01963 (2014).
4. R. Henderson, *Q. Rev. Biophys.* **28**, 171–193 (1995).
5. P. B. Rosenthal, R. Henderson, *J. Mol. Biol.* **333**, 721–745 (2003).
6. R. Henderson, G. McMullan, *Microscopy* **62**, 43–50 (2013).
7. A. Faruqi, R. Henderson, M. Pryddetch, P. Allport, A. Evans, *Nucl. Instrum. Methods* **546**, 170–175 (2005).
8. R. Henderson, R. M. Glaeser, *Ultramicroscopy* **16**, 139–150 (1985).
9. E. R. Wright, C. V. Iancu, W. F. Tivol, G. J. Jensen, *J. Struct. Biol.* **153**, 241–252 (2006).
10. R. M. Glaeser, G. McMullan, A. R. Faruqi, R. Henderson, *Ultramicroscopy* **111**, 90–100 (2011).
11. A. F. Brilot *et al.*, *J. Struct. Biol.* **177**, 630–637 (2012).
12. G. McMullan, S. Chen, R. Henderson, A. R. Faruqi, *Ultramicroscopy* **109**, 1126–1143 (2009).
13. G. McMullan, A. R. Faruqi, D. Clare, R. Henderson, *Ultramicroscopy* **147**, 156–163 (2014).
14. M. G. Campbell *et al.*, *Structure* **20**, 1823–1828 (2012).
15. X.-C. Bai, I. S. Fernandez, G. McMullan, S. H. Scheres, *eLife* **2**, e00461 (2013).
16. X. Li *et al.*, *Nat. Methods* **10**, 584–590 (2013).
17. C. J. Russo, L. A. Passmore, *Nat. Methods* **11**, 649–652 (2014).
18. D. Rhinow, W. Kühlbrandt, *Ultramicroscopy* **108**, 698–705 (2008).
19. C. Yoshioka, B. Carragher, C. S. Potter, *Microsc. Microanal.* **16**, 43–53 (2010).
20. C. J. Russo, L. A. Passmore, *J. Struct. Biol.* **187**, 112–118 (2014).
21. S. H. Banyard, D. K. Stammers, P. M. Harrison, *Nature* **271**, 282–284 (1978).
22. R. R. Crichton, J.-P. Declercq, *Biochim. Biophys. Acta* **1800**, 706–718 (2010).
23. W. H. Massover, *Micron* **24**, 389–437 (1993).
24. S. H. W. Scheres, *J. Struct. Biol.* **180**, 519–530 (2012).
25. N. de Val, J.-P. Declercq, C. K. Lim, R. R. Crichton, *J. Inorg. Biochem.* **112**, 77–84 (2012).
26. W. Baumeister, M. Hahn, J. Seredynski, L. M. Herbertz, *Ultramicroscopy* **1**, 377–382 (1976).
27. H. Stark, F. Zemlin, C. Boettcher, *Ultramicroscopy* **63**, 75–79 (1996).

ACKNOWLEDGMENTS

We thank R. Henderson for guidance and advice throughout this project; D. Mills and W. Kühlbrandt for the use of the Polara electron microscope at the Max Planck Institute of Biophysics, Frankfurt; I. S. Fernandez and the V. Ramakrishnan lab for the gift of ribosomes; G. McMullan, S. Chen, C. Savva, J. Grimmett, T. Darling, and M. Skehel for technical assistance; our colleagues at the Laboratory of Molecular Biology—especially S. Scheres, G. Murshudov, and R. A. Crowther—for many helpful discussions; and R. A. Crowther, V. Ramakrishnan, and E. Rajendra for a critical reading of the manuscript. This work was supported by the European Research Council (ERC) under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement no. 261151 and MRC grant U105192715. C.J.R. and L.A.P. are inventors on a patent application filed by the MRC on the gold substrates. The coordinates of the refined apoferritin model are deposited in the Protein Data Bank under accession code 4v1w and the EM density map is deposited in the Electron Microscopy Database under accession code 2788.

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/346/6215/1377/suppl/DC1
Materials and Methods
Figs. S1 to S10
Table S1
References (*28*–*40*)
Movie S1

4 August 2014; accepted 13 November 2014
10.1126/science.1259530

HIV ANTIBODIES

Antigen modification regulates competition of broad and narrow neutralizing HIV antibodies

Andrew T. McGuire,[1]*† Anita M. Dreyer,[1]*‡ Sara Carbonetti,[1] Adriana Lippy,[1] Jolene Glenn,[1]† Johannes F. Scheid,[2] Hugo Mouquet,[3] Leonidas Stamatatos[1,4]†§

Some HIV-infected individuals develop broadly neutralizing antibodies (bNAbs), whereas most develop antibodies that neutralize only a narrow range of viruses (nNAbs). bNAbs, but not nNAbs, protect animals from experimental infection and are likely a key component of an effective vaccine. nNAbs and bNAbs target the same regions of the viral envelope glycoprotein (Env), but for reasons that remain unclear only nNAbs are elicited by Env immunization. We show that in contrast to germline-reverted (gl) bNAbs, glnNAbs recognized diverse recombinant Envs. Moreover, owing to binding affinity differences, nNAb B cell progenitors had an advantage in becoming activated and internalizing Env compared with bNAb B cell progenitors. We then identified an Env modification strategy that minimized the activation of nNAb B cells targeting epitopes that overlap those of bNAbs.

During HIV-1 infection, the viral envelope glycoprotein (Env) elicits a polyclonal antibody response that targets diverse epitopes (*1*, *2*). Antibodies that display narrow breadth of neutralization (narrow neutralizing antibodies; nNAbs) develop during the first months of infection whereas those capable of neutralizing heterologous viruses (broadly neutralizing antibodies; bNAbs) develop several years later in ~10 to 30% of HIV-1–positive individuals (*3*). bNAbs isolated from HIV-1–infected patients are more protective than nNAbs in experimental HIV-1/SHIV infection (*4*) and will likely be a key component of an effective HIV-1 vaccine. Even though nNAbs and bNAbs target the same regions of Env (*2*, *5*–*7*), recombinant Env (rEnv) immunogens are poorly recognized by germline-reverted (gl) bNAbs (glbNAbs) and their corresponding B cell receptors (BCRs) (*5*, *8*–*21*), suggesting that the lack of bNAb generation during immunization may be due to inefficient stimulation of naïve bNAb BCR progenitors (*17*, *20*). In contrast, little is known about the recognition of rEnv by the naïve BCR progenitors of nNAbs. Understanding why B cell responses against nNAb epitopes dominate over those targeted by bNAbs in the context of rEnv immunization will inform on basic immunological mechanisms of epitope competition and provide new information relevant to the development of an effective HIV-1 vaccine.

Here we investigated whether glnNAbs from distinct clonal lineages that targeted the CD4-binding site (BS) and V3 regions of Env (*2*) also display minimal rEnv recognition. Amino acid differences between the mutated and gl sequences of nNAbs range from 2.4 to 7.3% for the heavy chains and 2.7 to 5.6% for the light chains for the nNAb CD4-BS antibodies (table S1 and fig. S1). In contrast, prototypic CD4-BS bNAbs, VRC01 (33.9% heavy, 23% light), NIH45-46 (a clonal relative of VRC01; 39.8% heavy, 26.1% light), b12 (21% heavy, 19% light), 8ANC131 (33% heavy, 24% light), and CH103 (12.7% heavy, 10% light) are more mutated (*5*, *8*, *16*, *22*). The anti-V3 nNAbs are more mutated (11.6 to 21.6% heavy, 9.7 to 13.8% light) than the anti-CD4-BS nNAbs.

In contrast to the anti-CD4-BS glbNAbs, which do not bind rEnv (*5*, *8*, *16*, *17*, *20*) (table S2), glnNAbs displayed broad Env recognition (from 51 to 100%) (table S2). The binding affinities of the glnNAbs were generally weaker than those of the corresponding mutated antibodies, owing to increased off rates in most cases (fig. S2).Whereas the glVRC01 class bNAbs were unable to neutralize any of the

[1]Seattle Biomedical Research Institute, Seattle, WA 98109, USA. [2]Laboratory of Molecular Immunology, The Rockefeller University, New York, NY 10065, USA. [3]Laboratory of Humoral Response to Pathogens, Department of Immunology, Institut Pasteur and CNRS-URA 1961, 75015 Paris, France. [4]Department of Global Health, University of Washington, Seattle, WA 98109, USA.
*These authors contributed equally to this work. †Present address: Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA. ‡Present address: GlycoVaxyn, Grabenstrasse 3, CH-8952 Schlieren, Switzerland. §Corresponding author. E-mail: lstamata@fhcrc.org

viruses tested, three of the five glnNAbs exhibited neutralizing activity against tier 1 viruses (table S3). Overall, we conclude that the glnNAbs and glbNAbs recognize the CD4-BS on soluble and virion-associated Env differently (*23*, *24*). Two of the three anti-V3 glnNAbs displayed neutralizing activity against several tier 1 viruses (table S3).

We next investigated whether B cells stably expressing glnNAb and glVRC01-class BCRs (fig. S3) could become activated by (Fig. 1A) and internalize (Fig. 1B) rEnv derived from clades A, B, and C. As previously reported, none of the rEnvs tested activated glVRC01-class B cells (*17*, *20*); however, they did activate glnNAb B cells targeting either the CD4-BS or V3 (Fig. 1A). Similarly, glnNAb B cells readily internalized diverse rEnvs, whereas glVRC01 class B cells did not (Fig. 1B). Combined, the above results indicate that rEnv immunogens can activate naïve nNAb B cells but not naïve VRC01-class B cells.

We recently reported that the disruption of three N-linked glycosylation sites (NLGS)—N276D, N460D, and N463D—on the clade C 426c rEnv (herein called 426c.NLGS.TM) confers binding to and activation of glNIH45-46 and glVRC01 B cells (*17*). In contrast, wild-type (WT) 426c Env is recognized by only the glnNAbs (table S2). gl1-154, gl1-695, gl1-732, and gl4-341 nNAbs inhibited binding of glNIH45-46 to 426c.NLGS.TM (fig. S5A), an indication that the epitopes of the anti-CD4-BS glnNAbs used here overlap those of the VRC01-class glbNAbs. The anti-V3 gl2-59 monoclonal antibody (mAb) caused a modest decrease

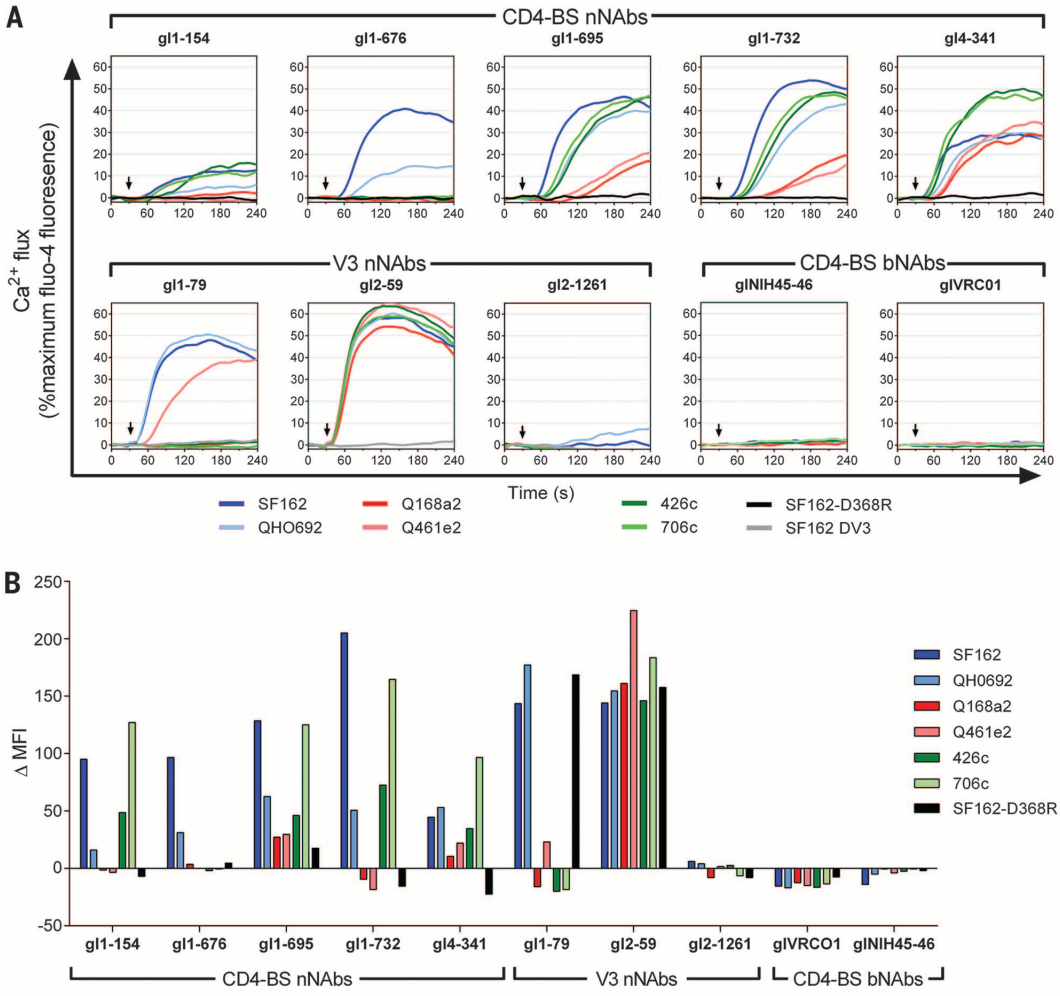


Fig. 1. Activation by and internalization of Env by glnNAb and glVRC01-class B cells. (**A**) Calcium flux in B cells expressing the glBCRs of CD4-BS specific nNAbs (1-154, 1-676, 1-695, 1-732, 4-341), CD4-BS specific bNAbs (NIH45-46 and VRC01), or V3-specific nNAbs (1-79, 2-59, 2-1261) challenged with the indicated Env proteins at a 1 μM final concentration. Black arrow indicates time of Env addition. (**B**) Quantitation of Env internalization in the above B cell lines. Bars represent the mean of two independent experiments (*31*). Clade A, B, and C rEnvs are colored shades of red, blue and green respectively in (A) and (B).

Table 1. Binding kinetics of germline reverted antibodies to trimeric 426c Env gp140 variants measured by BLI. NA: binding was undetectable.

	Ab	426c NLGS.TM trimeric gp140			426c NLGS.TM trimeric ΔV1Δ2Δ3		
		Apparent K_A (1/M)	k_{on}(1/Ms)	K_{off}(1/s)	Apparent K_A (1/M)	K_{on}(1/Ms)	K_{off}(1/s)
CD4-BS	glPt4-341	2.85×10^8	2.96×10^4	1.04×10^{-4}	1.77×10^8	1.57×10^4	8.85×10^{-5}
	glPt1-732	6.30×10^8	7.48×10^4	1.19×10^{-4}	2.65×10^8	9.71×10^3	3.66×10^{-5}
	glPt 1-154	4.12×10^7	1.61×10^4	3.91×10^{-4}	3.31×10^6	7.73×10^3	5.88×10^{-4}
	glPt 1-695	6.77×10^7	9.89×10^3	1.46×10^{-4}	2.24×10^7	4.66×10^3	2.08×10^{-4}
	glPt 1-676	NA	NA	NA	NA	NA	NA
	glNIH45-46	4.04×10^7	9.87×10^3	2.44×10^{-4}	3.52×10^8	1.40×10^4	3.98×10^{-5}
	glVRC01	2.67×10^8	5.51×10^3	2.06×10^{-4}	2.64×10^8	9.07×10^3	3.43×10^{-5}
V3	glPt1-79	NA	NA	NA	NA	NA	NA
	glPt2-59	1.19×10^{11}	1.85×10^5	1.55×10^{-6}	NA	NA	NA
	glPt2-1261	NA	NA	NA	NA	NA	NA

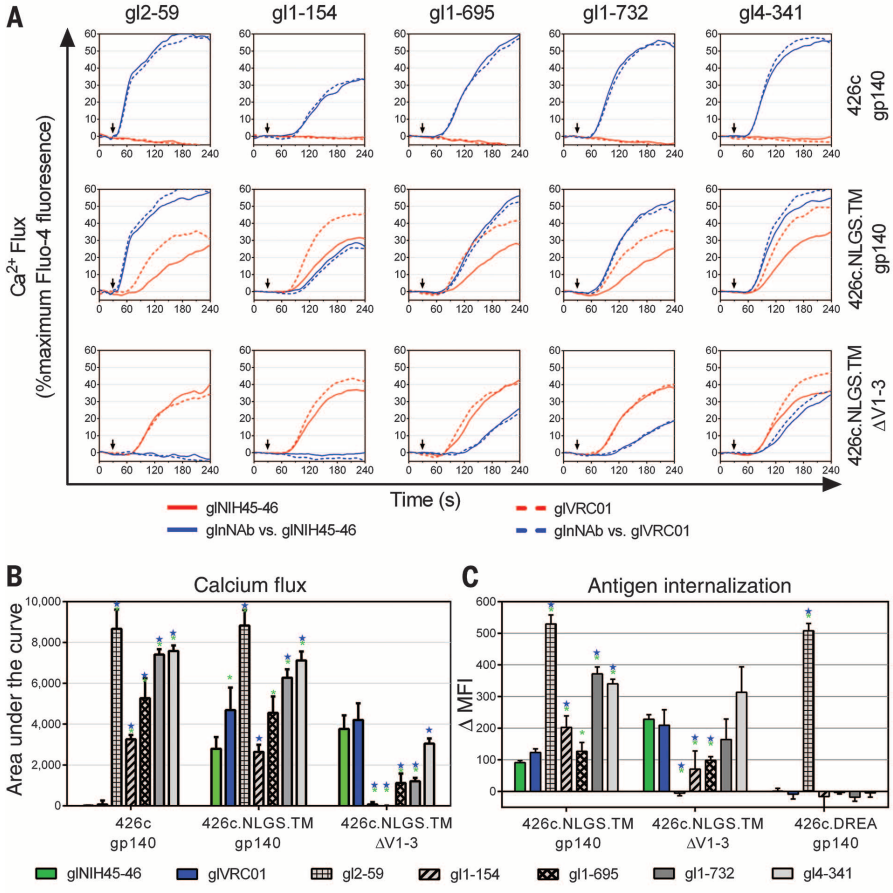


Fig. 2. Activation by and internalization of Env by glnNAb and glVRC01-class B cells in response to modified Env immunogens. (**A**) glNIH45-46 or glVRC01 B cells were mixed with the indicated glnNAb B cells. The pooled B cells were then challenged with WT 426c (top row), 426c.NLGS.TM (middle row), or 426c.NLGS.TM.ΔV1-3 (bottom row) at a 1 μM final concentration. Calcium flux was monitored concurrently in both B cell lines with the strategy depicted in fig. S6. Black arrow indicates time of Env addition. (**B**) Area under the calcium flux curve determined from duplicate experiments in (A). Error bars represent SD from the mean of 10 (glVRC01-class) or 4 replicates (other B cell lines). Green asterisks and blue stars over the bars indicate significant differences ($P < 0.05$) from glNIH45-46 and glVRC01 B cells, respectively, for each protein using a two-tailed Student's *t* test. (**C**) Quantification of internalization of the indicated Envs by glVRC01 class and glnNAb B cells. Error bars represent SD from the mean of two independent experiments each performed in duplicate ($n = 4$). Green astrices and blue stars represent significant differences ($P < 0.05$) from glNIH45-46 and glVRC01, respectively, for each protein using a two-tailed Student's *t* test.

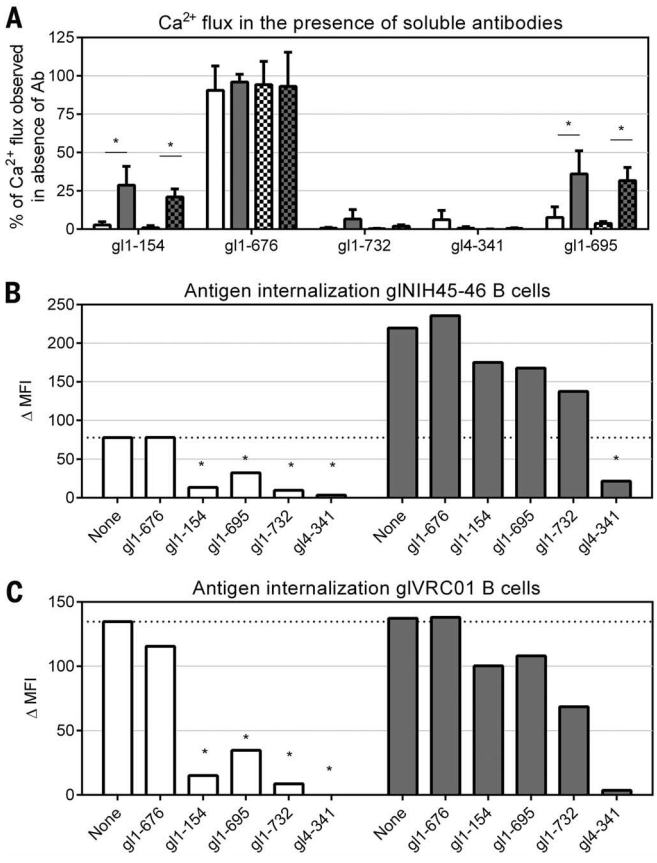


Fig. 3. B cell activation and antigen uptake of Env by glVRC01-class B cells in the presence of soluble anti-HIV antibodies. (**A**) Calcium flux was monitored in glNIH45-46 (solid bars) or glVRC01 (patterned bars) B cells incubated with 426c.NLGS.TM (white bars) or with 426c.NLGS.TM.ΔV1-3 (gray bars) in the presence of equimolar concentrations of the indicated antibodies. Bars represent mean and SD of the area under the calcium flux response curve relative to that of Env in the absence of soluble mAb from three independent experiments. Asterisks represent a significant difference ($P < 0.05$) in the calcium flux response between 426c.NLGS.TM and 426c.NLGS.TM.ΔV1-3 using a two-tailed Student's *t* test. Env internalization by glNIH45-46 (**B**) or glVRC01 (**C**) B cells incubated with 426c.NLGS.TM (white bars) or 426c.NLGS.TM.ΔV1-3 (gray bars) in the absence or presence of equimolar concentrations of the indicated nNAbs. Bars represent the mean of two independent experiments. Asterisk denotes significant difference ($P < 0.05$) in fluorescence compared to cells incubated with Env alone using a two-tailed Student's *t* test in (B) and (C). Dotted line corresponds to mean 426c.NLGS.TM internalization by B cells in the absence of Ab in (B) and (C).

in the binding of glNIH45-46 (fig. S5A). The gl1-676, gl1-79, and gl2-1261 antibodies, which do not bind 426c.NLGS.TM (Table 1), had no effect on the binding of glNIH45-46 (fig. S5A).

In germinal centers (GCs), B cells expressing higher-affinity BCRs selectively expand, whereas lower-affinity B cells are eliminated (*25*). The anti-CD4-BS gl1-154, gl1-695, gl1-732, and gl4-341, and the anti-V3 gl2-59 bound 426c.NLGS.TM with a higher apparent affinity than glNIH45-46 (Table 1). glVRC01 bound more strongly to 426c.NLGS.TM than gl1-154 and gl1-695, but had a slightly lower apparent affinity than the gl4-341 and gl1-732 mAbs. The gl2-59 V3–directed mAb bound more strongly to 426c.NLGS.TM than any of the glCD4-BS Abs, consistent with the higher propensity of gl2-59 B cells to be activated and take up rEnv (Fig. 1). We anticipate that upon immunization with the 426c.NLGS.TM, naïve 2-59–like B cells, or B cells that target immunodominant epitopes on the variable regions of Env (V1, V2 and V3), would have an advantage over naïve VRC01-class B cells in becoming activated, taking up antigen, and obtaining T cell help.

To investigate the effect that differences in binding affinity have on the relative activation of B cells expressing different anti-Env BCRs, we developed an assay that allows for the concurrent monitoring of B cell activation in distinct B cell populations (fig. S6). As expected, WT 426c activated glnNAb B cells (Fig. 2A top row, and Fig. 2B), but not glVRC01-class B cells (Fig. 2A top row, and Fig. 2B). 426c.NLGS.TM activated gl VRC01-class B cells, but with the exception of gl1-154, it activated glnNAb B cells more strongly (Fig. 2A, middle row). 426c.NLGS.TM activated gl1-695 more strongly than glNIH45-46 and comparably to glVRC01 (Fig. 2A, middle row, and Fig. 2B).

Because 426c.NLGS.TM showed the strongest activation of gl2-59 B cells, we aimed to abolish activation of B cells targeting the V1, V2, and V3 regions of 426c.NLGS.TM, by deleting these regions to create 426c.NLGS.TM.ΔV1-3. As expected, 426c.NLGS.TM.ΔV1-3 did not bind to gl2-59 mAb (Table 1) and did not activate gl2-59 B cells (Fig. 2, A and B). With the exception of glNIH45-46, deletion of the variable loops reduced the apparent binding affinities of the anti-CD4-BS antibodies tested (Table 1). Consistent with these changes in binding, 426c.NLGS.TM.ΔV1-3 activated glNIH45-46 B cells more robustly than 426c.NLGS.TM (Fig. 2, A and B). The relative activation of all but one anti-CD4-BS glnNAb B cell (gl4-341) by 426c.NLGS.TM.ΔV1-3 was less than that observed for glVRC01-class B cells (Fig. 2, A and B). gl4-341 B cells were activated comparably to glVRC01 class B cells, although the activation was reduced relative to that observed with WT 426c and 426c.NLGS.TM (Fig. 2, A and B). At lower concentrations of 426c.NLGS.TM.ΔV1-3, only glVRC01-class and gl4-341 B cells were activated (fig. S7). To represent a more physiological B cell repertoire, we challenged glVRC01-class B cells mixed with an excess of pooled glnNAb, and non-Env reactive B cells (fig. S8). WT 426c only activated nNAb B cells, and although 426c.NLGS.TM could activate the minority glVRC01-class B cells, a preferential activation of the nNAb B cell pool was observed. In contrast, 426c. NLGS.TM.ΔV1-3 activated glVRC01-class B cells preferentially at higher concentrations and exclusively at lower concentrations (fig. S8). When the internalization of rEnv by B cells was quantitated, 426c NLGS.TM was preferentially internalized by nNAb B cells, whereas 426c NLGS.TM.ΔV1-3 was preferentially internalized by glVRC01-class and gl4-341 B cells (Fig. 2C). The differential impact of removing the variable Env regions on the binding affinities, B cell activation, and antigen internalization for the bNAbs versus nNAbs to the CD4-BS is likely due to differences in the angle of approach to the CD4-BS by these two types of antibodies (*16*, *22*, *24*, *26–29*). Additionally, the nNAbs may make contacts with portions of the variable loops.

Antibodies can enter GCs and inhibit BCR binding to an immunogen unless the affinity of the BCR for the immunogen is higher than for the soluble antibody (*30*). Through this feedback mechanism, higher-affinity antibodies hinder the ability of low-affinity, antigen-specific B cells to receive T cell help and thus select for higher-affinity B cell clones (*30*). Because the apparent affinities of glVRC01-class bNAbs for 426c.NLGS.TM ΔV1-3 are higher than those for the glnNAbs studied here (Table 1), we expect that such a feedback mechanism will provide a selective advantage to glVRC01-class B cell responses if this rEnv is used as an immunogen. Inversely, owing to the higher apparent affinities of the glnNAbs for the 426c.NLGS.TM relative to the glVRC01 class bNAbs (Table 1), we would expect this immunogen to drive the development of nNAb responses through the same antibody-feedback mechanism. Indeed, the binding of anti-CD4-BS nNAbs to 426c.NLGS.TM reduced the activation of glVRC01-class B cells by that rEnv (Fig. 3A and fig. S9). In contrast, activation of glVRC01 class B cells by 426c.NLGS.TM.ΔV1-3 rEnv was observed even in the presence of the gl1-154, and gl1-695 Abs (Fig. 3A and fig. S9). Thus, removal of the V1, V2, and V3 regions reduces the ability of some (but not all) anti-CD4 nNAbs to block rEnv activation of glVRC01 class B cells.

Similar observations were made when the internalization of these two Envs by glVRC01 class B cells was determined in the absence and presence of glnNAbs (Fig. 3, B and C). glNIH45-46 B cells more readily internalized 426c.NLGS.TM.ΔV1-3 compared to 426c.NLGS.TM, even in the presence of inhibitory antibodies, with the exception of the gl4-341 mAb (Fig. 3B, compare white and gray bars). Overall, we conclude that the presence of anti-CD4-BS glnNAbs can inhibit the activation of, and the antigen internalization by, glVRC01-class B cells in response to the 426c.NLGS.TM, but that these effects are less pronounced when the variable regions V1, V2, and V3 are removed from this rEnv. Thus, selecting which immunogen is used during the priming phase of immunization (e.g., 426c.NLGS.TM versus 426c.NLGS.TM.ΔV1-3) will be critically important to the eventual elicitation of anti-CD4-BS bNAbs.

In sum, our study provides a mechanistic explanation as to why rEnv immunogens elicit narrow rather than broadly neutralizing antibody responses to the CD4-BS of Env. This is primarily due to the broad Env-recognition properties of glnNAb BCRs, which contrasts with the inefficiency by which bNAb BCRs recognize rEnv. Notably, we demonstrate that rational immunogen modifications can reduce (and in certain cases eliminate) the activation of naïve B cells that give rise to such nNAbs, while promoting the activation of naïve B cells that give rise to glVRC01-class bNAbs, even when the epitope of these antibodies substantially overlap. As such, our results are relevant to efforts to elicit protective antibody responses against not only HIV-1 but potentially other pathogens as well.

REFERENCES AND NOTES

1. H. Mouquet *et al*., *PLOS ONE* **6**, e24078 (2011).
2. J. F. Scheid *et al*., *Nature* **458**, 636–640 (2009).
3. L. Stamatatos, L. Morris, D. R. Burton, J. R. Mascola, *Nat. Med.* **15**, 866–870 (2009).
4. F. Klein *et al*., *Science* **341**, 1199–1204 (2013).
5. J. F. Scheid *et al*., *Science* **333**, 1633–1637 (2011).
6. J. E. Robinson *et al*., *J. Virol.* **84**, 3443–3453 (2010).
7. A. Changela *et al*., *J. Virol.* **85**, 2524–2535 (2011).
8. S. Hoot *et al*., *PLOS Pathog.* **9**, e1003106 (2013).
9. X. Xiao, W. Chen, Y. Feng, D. S. Dimitrov, *Viruses* **1**, 802–817 (2009).
10. X. Xiao *et al*., *Biochem. Biophys. Res. Commun.* **390**, 404–409 (2009).
11. M. Huber *et al*., *J. Virol.* **84**, 10700–10707 (2010).
12. H. Mouquet *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **109**, E3268–E3277 (2012).
13. D. Sok *et al*., *PLOS Pathog.* **9**, e1003754 (2013).
14. M. Bonsignori *et al*., *J. Virol.* **85**, 9998–10009 (2011).
15. B. J. Ma *et al*., *PLOS Pathog.* **7**, e1002200 (2011).
16. T. Zhou *et al*., *Science* **329**, 811–817 (2010).
17. A. T. McGuire *et al*., *J. Exp. Med.* **210**, 655–663 (2013).
18. T. Ota *et al*., *J. Immunol.* **189**, 4816–4824 (2012).
19. K. J. Doores *et al*., *J. Virol.* **87**, 2234–2241 (2013).
20. J. Jardine *et al*., *Science* **340**, 711–716 (2013).
21. A. T. McGuire, J. A. Glenn, A. Lippy, L. Stamatatos, R. W. Doms, *J. Virol.* **88**, 2645–2657 (2014).
22. H.-X. Liao *et al*., *Nature* **496**, 469–476 (2013).
23. R. W. Sanders *et al*., *PLOS Pathog.* **9**, e1003618 (2013).
24. D. Lyumkis *et al*., *Science* **342**, 1484–1490 (2013).
25. G. D. Victora, M. C. Nussenzweig, *Annu. Rev. Immunol.* **30**, 429–457 (2012).
26. T. Zhou *et al*., *Nature* **445**, 732–737 (2007).
27. L. Chen *et al*., *Science* **326**, 1123–1127 (2009).
28. R. Diskin *et al*., *Science* **334**, 1289–1293 (2011).
29. T. Zhou *et al*., *Immunity* **39**, 245–258 (2013).
30. Y. Zhang *et al*., *J. Exp. Med.* **210**, 457–464 (2013).
31. Cells expressing the g1-732 BCR did not internalize the clade A Q168a2 and Q461e2 rEnvs that activated B cells expressing that BCR. This discrepancy can be attributed to pHrodo labeling of these two rEnvs, which reduced the binding of the gl1-732 mAb (fig. S4).

ACKNOWLEDGMENTS

We thank M. Nussenzweig for providing the nNAb antibody expression plasmids, J. Mascola and X. Wu for providing the IgG1 glVRC01 antibody, D. Rawlings for providing the pRRL expression plasmids, and L. Scharf and P. Bjorkman for technical assistance. The data presented in this paper are found in the main paper and in the supplementary materials. Accession numbers for the mutated and germline-reverted nNAbs used in this study can be found at GenBank: KM595240-KM595271. This work was supported by NIH grants P01 AI094419-01 and U19 19AI109632-01 (L.S.), Canadian Institutes of Health Research fellowship (A.T.M), and Swiss National Science Foundation fellowship PBBSP3_144245 and P300P3_151140 (A.M.D.).

SUPPLEMENTARY MATERIALS

www.sciencemag.org/content/346/6215/1380/suppl/DC1
Materials and Methods
Author Contributions
Figs. S1 to S9
Tables S1 to S3
References (*32–43*)

25 July 2014; accepted 13 November 2014
10.1126/science.1259206

Quantitative Western Blot System

The Amersham WB system is a fully integrated system for quantitative SDS-PAGE and Western blotting of proteins using fluorescence detection. The system is designed to drive down the variability seen with conventional Western blotting. With the Amersham WB system, every stage of the Western blotting process, including electrophoresis, transfer, probing, and scanning, is standardized and monitored—resulting in consistent, quantitative protein analysis. Western blotting requires a complex combination of analytical knowledge, extensive manual handling, dexterity, and strict adherence to protocols in order to achieve reliable results. Failure rates, however, can be as high as 25% and data variations between different users can be as high as 35%–40%. To address this shortfall in consistency, GE Healthcare Life Sciences has developed the Amersham WB system, with a standardized traditional Western blotting workflow that typically achieves a variability of less than 10% between users. The system's consistency and reproducibility delivers robust quantitative data.

GE Healthcare Life Sciences
For info: +44-(0)-0800-032-9201
www.gehealthcare.com

Protein A Kit

The Gyrolab Protein A Kit is designed for use on the Gyrolab xP workstation. During purification of biotherapeutics, Protein A can leach from the chromatography support and co-elute with the therapeutic antibody product. The contamination complicates subsequent analysis using immunoassays. In addition, Protein A can elicit an immune response. It is therefore essential to accurately quantify Protein A during the purification process, and prior to product release. The new Protein A kit, designed and validated for use with Gyros nanoliter-scale immunoassay platform, includes automated acid dissociation of Protein A and IgG during sample preparation, to ensure contamination is accurately assessed. Gyros has incorporated proprietary antibodies to ensure high specificity and sensitivity. The easy-to-use kit quantifies native Protein A, recombinant Protein A variants, and MabSelect SuRe ligand, and has fewer manual steps than other standard enzyme-linked immunosorbent assay kits, reducing hands-on time and risk for error.

Gyros
For info: +46-(0)-18-56-63-00
www.gyros.com



Western Blot Imaging System

Until now, digital imaging systems have failed to deliver the sensitivity and resolution of film. The new ChemiDoc Touch System, however, allows detection of faint bands missed by film and produces publication-quality images. The ChemiDoc Touch outperforms film in other ways as well. When using film to image abundant proteins, strong bands quickly saturate and become unquantifiable. Saturated bands can also obscure the signal from adjacent faint bands, making Western blotting with film challenging. The ChemiDoc Touch addresses these issues through its wide dynamic range, which permits easy and reliable quantification even of highly abundant proteins, and through the instrument's ability to optimize exposure for each protein of interest. Capturing images with the ChemiDoc Touch is easy. The ChemiDoc Touch offers a smooth, intuitive user experience that makes capturing, reviewing, selecting, and exporting images efficient and straightforward. The imager also allows stain-free imaging, a technology exclusive to Bio-Rad.

Bio-Rad Laboratories
For info: 800-424-6723
www.bio-rad.com/chemidoctouchpr

Phosphorylation Detection Assays

AlphaLISA SureFire Ultra assays are designed to reduce assay time and to improve the protein phosphorylation detection process. With high sensitivity and wide dynamic range, AlphaLISA SureFire Ultra assays are quick and easy to run with a minimal amount of sample, significantly reducing time and expense. The AlphaLISA SureFire Ultra assays use TGR BioSciences' CaptSure technology to eliminate antibody interference in the therapeutic antibody screening process. AlphaLISA SureFire Ultra sensitivity and dynamic range also make it ideal for other types of screening, such as small molecule modulation, detection of low and basal levels of phosphorylation, and detection of phosphorylation in tissue lysates. The assays also leverage PerkinElmer's AlphaLISA technology that has better spectral characteristics related to hemoglobin interference, which can otherwise hinder the accuracy of measurements. AlphaLISA SureFire Ultra generates results in less than two and a half hours with only one or two incubation steps and no washing.

PerkinElmer
For info: 800-762-4000
www.perkinelmer.com/surefire

Remote Access Imaging System

The vibrant new T:Genius imaging system allows high-performance, walkaway imaging of gels and blots on a smartphone or tablet. Using the T:Genius' "StatusLink" feature, researchers can stay updated and even share image results with colleagues, no matter whether they are down the hall or in another country. This is a great time saver when, for instance, scientists are developing a long-exposure chemi blot and want to see how it's doing while having lunch or busying themselves with something else in the lab. The new T:Genius is both stylish and practical, comes in a choice of exciting colors, and features a sensitive camera that will give high-res images. With the powerful GeneSys image capture software, researchers can set up their imaging protocols, click, walk away, and see accurate DNA or protein results later in the lab or on their tablet or phone.

Syngene
For info: +44-(0)-1223-727123
www.syngene.co.uk/t-genius

Electronically submit your new product description or product literature information! Go to www.sciencemag.org/products/newproducts.dtl for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.

Downloaded from www.sciencemag.org on December 14, 2014

ScienceCareers online @sciencecareers.org



There's only one Science

Science Careers Advertising

For full advertising details, go to ScienceCareers.org and click For Employers, or call one of our representatives.

Tracy Holmes
Worldwide Associate Director
Science Careers
Phone: +44 (0) 1223 326525

THE AMERICAS

E-mail: advertise@sciencecareers.org
Fax: 202 289 6742

Tina Burks
Phone: 202 326 6577

Nancy Toema
Phone: 202 326 6578

Marci Gallun
Sales Administrator
Phone: 202 326 6582

Online Job Posting Questions
Phone: 202 312 6375

EUROPE / INDIA / AUSTRALIA / NEW ZEALAND / REST OF WORLD

E-mail: ads@science-int.co.uk
Fax: +44 (0) 1223 326532

Axel Gesatzki
Phone: +44 (0) 1223 326529

Sarah Lelarge
Phone: +44 (0) 1223 326527

Kelly Grace
Phone: +44 (0) 1223 326528

JAPAN

Katsuyoshi Fukamizu (Tokyo)
E-mail: kfukamizu@aaas.org
Phone: +81 3 3219 5777

Hiroyuki Mashiki (Kyoto)
E-mail: hmashiki@aaas.org
Phone: +81 75 823 1109

CHINA / KOREA / SINGAPORE / TAIWAN / THAILAND

Ruolei Wu
Phone: +86 186 0082 9345
E-mail: rwu@aaas.org

All ads submitted for publication must comply with applicable U.S. and non-U.S. laws. *Science* reserves the right to refuse any advertisement at its sole discretion for any reason, including without limitation for offensive language or inappropriate content, and all advertising is subject to publisher approval. *Science* encourages our readers to alert us to any ads that they feel may be discriminatory or offensive.

ScienceCareers
FROM THE JOURNAL SCIENCE AAAS

ScienceCareers.org

ScienceCareers

Cernet

"《科学》职业" 已经与Cernet/赛尔互联开展合作。中国大陆的高校可以直接联系Cernet/赛尔互联进行国际人才招聘。



请访问
Sciencecareers.org/CER
点得联系信息。

招募学术精英，《科学》是您的不二之选 Science



Giant Pandas

Exploring the Frontiers of Science in Western China

China's remarkable economic rise is also propelling advances across the sciences, and these intertwined trends are transforming the western section of the country. To speed up this transformation, the Chinese Academy of Sciences (CAS), universities, and the central government are all creating award schemes aimed at attracting leading scholars trained abroad to take up positions in western China. Science centers across the region are ramping up efforts to make research teams more innovative and more international. This strategy extends to the spheres of genomics, medicine, geoscience, astronomy, physics, and conservation and is creating professional openings across the western sector of China. **By Kevin Holden**

China's evolution into an economic superpower is advancing simultaneously with a global rise in scientific research. Research funding distributed by the National Natural Science Foundation of China (NSFC) is expanding rapidly, in step with campaigns to attract Chinese scientists who have obtained advanced degrees abroad. The foundation, which was established with an annual budget of 80 million yuan (~US$13 million) in 1986, saw that figure balloon to 2 billion yuan in 2003 (~US$325,000), and to 23.8 billion yuan (~US$3.87 billion) in 2013, according to the group's annual report. One quarter of the foundation's annual funding is channeled into its Talent Training Program, which is aimed at cultivating leading professionals and research groups.

The head of the NSFC, **Yang Wei**, said in an interview with *Science* magazine in 2013 (scim.ag/1y4rFhc) that the foundation's funding is projected to continue increasing steeply in the future. "China has to transition from an economic powerhouse to a technological powerhouse and then to a scientific and cultural powerhouse," he said. "To achieve this goal, we will need many scientists, and we need to convince the government that it should provide more funding to the NSFC."

The NSFC president is just one of the key figures across China who are helping push the sciences forward, in part through the proliferation of awards aimed at enticing leading scholars—both Chinese expatriates and international scientists—to relocate to China. The drive to advance research has generated an array of opportunities for scientists—a growing number of which are spread out across the western Chinese provinces of Sichuan and Gansu.

Sichuan, in southwest China, is home to 80 million people. The region has seen a surge in research focused on restoring and preserving ecosystems (due to its expanding population and a massive 2008 earthquake) and on the life sciences, such as genomics and medicine. The capital city of Chengdu is dotted with ancient temples and teahouses, but also holds a cluster of some of China's newest universities, including the Chengdu University of Technology, a science and engineering school that hosts more than 32,000 students and over three dozen research institutes.

In northwestern China's Gansu Province—part of the ancient Silk Road, a 2,000-year-old trade route that linked China with Middle Eastern and Mediterranean cultures—scientific research extends as far as the Tibetan Plateau. Lanzhou, the capital, is surrounded by dramatic mountain peaks, and still has the feel of a timeworn oasis. Lanzhou University, which hosts nearly 20,000 undergraduates and more than 10,000 postgraduates, is a region-wide magnet for scholars and students in physics, chemistry, and information sciences.

Gansu and Sichuan each hosts a major spaceflight launch center, and both provinces aim to ramp up research and development of space technology as part of a wider push to accelerate progress across the sciences.

The eastern edge of the Tibetan Plateau cuts through both provinces, and Tibetan communities are sprinkled across its foothills. The region is a gateway for trekkers and researchers eager to explore the mountains and cultures of westernmost China.

A new "Golden Age" in southwestern China

"A sense of adventure" impelled **Nicholas Kersting**, who obtained a Ph.D. in particle physics from the University of California, Berkeley, to do his postdoctoral research at Sichuan University.

Kersting, who until recently was a professor at Sichuan University is now a senior associate at continued>

PHOTO: SHAHREEN/SHUTTERSTOCK.COM

Upcoming Features

Faculty—January 30, 2015 ■ *Postdocs—February 27, 2015* ■ *Regional Focus: Japan—March 27, 2015*

Princeton Consultants in New Jersey, says his interest in the region initially was sparked when he visited to give a research talk at the school's physics department. "I became very fond of the incredible cuisine and mountain scenery, and the culture felt electric," he explains. "I needed to explore that."

Looking ahead, he predicts: "Sichuan University will attract increasing numbers of Western scholars, partly due to the improving condition of the university and organization of the physics department."

"The ambiance of Chengdu was always addictive, but now the overall quality of life is accelerating far more than any of the many other cities I've been to in China," he explains. "I think we're entering its Golden Age."



"Earth, ecological, and environmental sciences are all on the rise."
–Kang Shichang

Ji Xiangdong, a professor in theoretical physics at the University of Maryland who also heads the Institute of Nuclear and Particle Physics, Astronomy, and Cosmology at Shanghai Jiao Tong University, says China's strengthening economic power is creating a widening web of major science projects and opportunities for scientists from across the world. This is helping reverse a decades-long brain drain that has seen many Chinese scientists who studied in the West stay in North America or Europe.

Ji also heads the PandaX Experiment, a search for dark matter being conducted from an underground laboratory, created by cutting a 17 km long tunnel through a mountain on the outskirts of Chengdu. Chinese and American scientists collaborating on the project are using a xenon-based detector to monitor potential collisions between xenon nucleons and weakly interacting massive particles, hypothesized candidates for dark matter particles. "What's really exciting is a plan to enlarge this new underground lab by about 20 times, creating space for more scientific experiments," he explains, making the city even more attractive for leading scientists from around the world.

The experiment is partially funded by the NSFC, whose outlays for research across the sciences are surging in tandem with China's high-speed GDP growth, he says, even as the central government increases special financing for science centers in western China.

The Chinese Academy of Sciences, meanwhile, has already attracted more than 2,000 scholars with advanced degrees from abroad with its award programs, and is launching a new Hundred Talents Plan to step up this strategy, says **Cao Jinghua**, deputy director-general of the CAS Bureau of International Cooperation.

CAS recently hosted a "Joint Scientific Space Mission Workshop" with the European Space Agency in Chengdu that brought together scientists from across Europe and China. "Space is one area that we are emphasizing for increasing international cooperation," he explains, and adds the workshop was staged in Chengdu to showcase the region's strengthening science centers.

Chengdu is tapping burgeoning state coffers to build up its universities and academies, part of a nationwide strategy aimed at using scientific knowledge to shape a better collective future, according to **Jiang Jianping**, a senior scientist at the CAS Chengdu Institute of Biology.

While the central government is investing more in Sichuan, it is also "making better policies to attract scholars, especially scientists, into the region," he explains. At the same time, he says, "The Chinese Academy of Sciences has a plan to attract more researchers from Europe and America."

The Chengdu Institute of Biology already hosts visiting scholars from Europe and will likely become increasingly international, explains **Dai Qiang**, an associate professor studying animal ecology and conservation biology. Its research centers in agricultural biotechnology, natural medicinal compounds, biopharmaceuticals, environmental microbiology, and ecology are all currently expanding, he adds.

The biology institute's National Engineering Research Center for Natural Medicines has teamed up with Chengdu Di'ao Pharmaceutical Group to research and produce medicines derived from natural compounds. **Ji Jianxin**, a professor in pharmacological chemistry who also heads research at Di'ao, explains that while the research center aims to import and develop new technologies to enhance pharmaceutical production across China, Di'ao—which had one traditional herbal medicinal product approved for sale in the Netherlands in 2012—ultimately will target the entire European Union and U.S. markets. Di'ao has already assembled a small circle of American- and European-educated scientists, and plans to attract even more scholars from around the world, he adds. Ji himself was granted a CAS Hundred Talents award after his postdoctoral research at Vanderbilt University in Nashville, Tennessee.

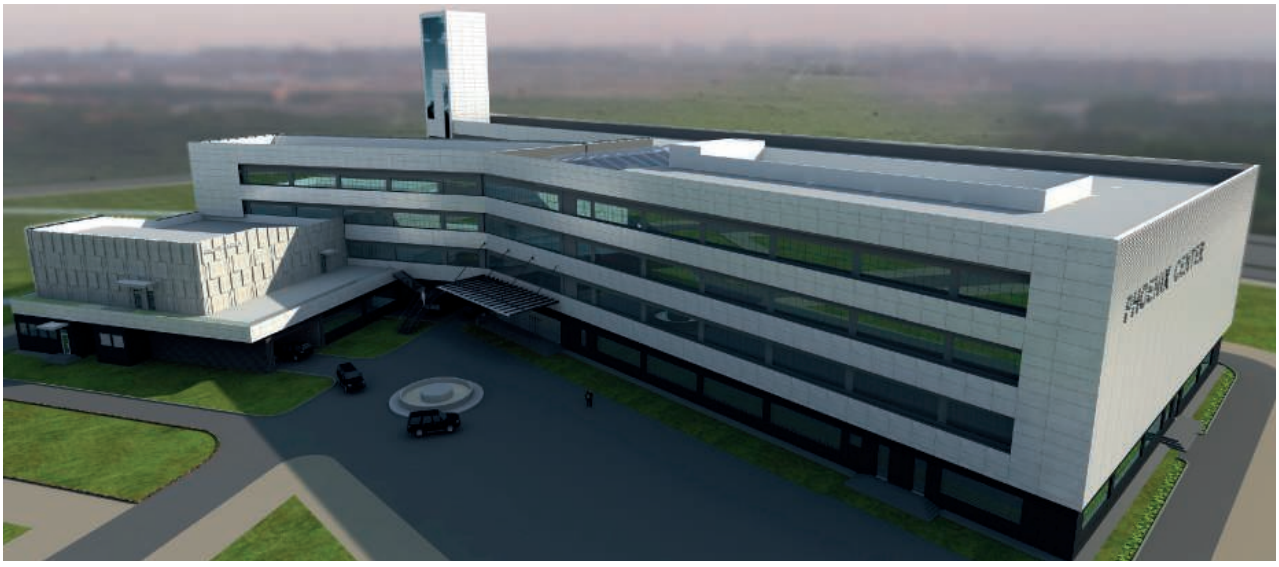
The CAS outpost in Chengdu also hosts a range of environmentally focused projects. **Wu Ning**, the director of the Chengdu Institute of Biology, investigates restoration of China's degraded mountain ecosystems, the impacts of the catastrophic 2008 earthquake on Sichuan's environment, and the effects of rangeland privatization on biodiversity across the Tibetan Plateau.

Some of the research positions being opened to Western-trained scientists can be both intensely continued>

OPEN POSITIONS IN PROTEOMICS AND TRANSLATIONAL MEDICINE

The National Center for Protein Sciences (the PHOENIX Center), Beijing, China

National Center for Protein Sciences



In the last decade, China has undergone phenomenal growth in proteomics research, encompassing both methodology development and bio-medical applications. The PHOENIX Center (Pilot Hub Of ENcyclopedicproteomIX) is finishing construction and will open its doors in the spring of 2015. It is located in Zhongguancun Life Science Park (ZLS Park) in the northwest suburb of Beijing, an important life science R&D hub in China. The PHOENIX Center will be a place that integrates proteomics and translational biology. We are seeking candidates to fill 30 or so faculty positions at all academic levels. We are also soliciting applicants for the positions of the Center Director and the Section Chief to lead the 5 divisions (Analytical Proteomics, Functional Proteomics, Translational Medicine, Bioinformatics, and Supporting Facilities) in the PHOENIX Center. These positions will begin in August 2015. The appointments come with generous start-up packages including competitive personal compensations. The starting annual salary is 300K RMB for positions at the assistant professor level.

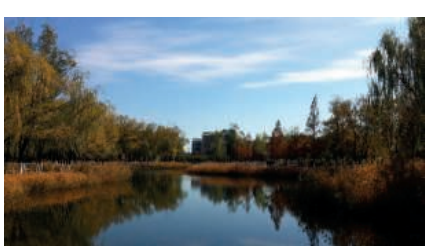
Please submit the following materials to phoenixcenter_hr@163.com, no later than February 28, 2015:

1. Cover letter
2. *Curriculum Vitae*
3. Research statement
4. Names and contact information of three to five references (three letters of recommendation are required, and the application is complete only when all three letters have been submitted)

Contact Information: Mr. Liang Zhao, BPRC, 33 Life Science Road, ChangPing, Beijing, 102206, China

Contact Email: phoenixcenter_hr@163.com







JP254841R

ScienceCareers online @sciencecareers.org
FOCUS ON CHINA


Zhang Wenping

Featured Participants

CAS Bureau of International Cooperation
scim.ag/1vZAZI9

Chengdu Di'ao Pharmaceutical Group
www.diao.com

Chengdu Institute of Biology, CAS
english.cib.cas.cn

Chengdu Research Base of Giant Panda Breeding
scim.ag/1v4R5sT

Institute of Modern Physics, CAS
scim.ag/1yvp1DY

Institute of Mountain Hazards and Environment, CAS
english.imde.cas.cn

Lanzhou University
www.lzu.edu.cn/notice/English/Introduction.htm

Princeton's Institute for Advanced Study
www.ias.edu

Shanghai Jiao Tong University
en.sjtu.edu.cn/

Sichuan University
www.scu.edu.cn/en/about/H6401index_1.htm

State Key Laboratory of Cryospheric Science CAS
scim.ag/1HS0LB6

exciting and have immense benefit to society, notes **Cui Peng**, vice director of the CAS Institute of Mountain Hazards and Environment in Chengdu. In May of 2008, in the immediate aftermath of magnitude 8.0 earthquake in Sichuan, Cui Peng led a team that examined extremely dangerous dams created by landslides and rock falls that threatened millions of lives across the seismic hazard zone. Their rapid assessment of the massive quake-induced transformation of the region's geography and their engineering solutions to these geohazards likely saved countless lives.

The quake also affected many of the mountainside refuges that are the last natural habitats of the giant panda; even some of the rebuilding projects that followed led to increased fragmentation of these panda reserves. This ever-looming threat impelled scientists at the Chengdu Research Base of Giant Panda Breeding to accelerate programs aimed at saving the panda.

This research base has attracted top scholars in biomedicine, reproductive biology, genomics, and conservation who have teamed up on studies aimed at pulling the panda back from the brink of extinction.

Professor **Dai Qiang** has been working with scientists at the panda research base to create computer simulations mapping the effects of habitat fragmentation or expansion on the collective gene pool of endangered species. Dai says that these studies could be used to help giant pandas and other wildlife facing the threat of extinction across Sichuan.

Scientists at this research base likewise have created a global web of collaborations with counterparts from organizations including the Smithsonian National Zoo, Zoo Atlanta, and the Madrid Zoo to test the panda's ability to adapt to new ecosystems across the continents. By sending panda couples to zoos across the world, Chinese guardians of the panda are allowing researchers at these outposts to experiment with varying breeding techniques, habitats, and even panda-focused child-rearing psychology studies. Breakthroughs and discoveries made via this globe-spanning panda network are channeled back to the Chengdu research base and applied to aid the panda's survival.

Other scientists at the Chengdu research base conduct periodic reviews of the genetic profiles of the captive and wild populations of the giant panda, and have proposed schemes to ensure gene flow across these groups to prevent genetic bottlenecks. Researchers have also joined up with international colleagues to assemble a draft sequence of the giant panda genome and to help determine whether the estimated 2,500 extant pandas feature enough genetic variability to prevent the destruction of the species.

The research base aims to recruit scholars worldwide to join the core team in southwest China, according to **Zhang Wenping**, a genomics expert at Chengdu's panda breeding center.

Scientific oases in northwestern China

An expanding arc of scientific research areas is attracting scholars to western China: "Earth, ecological, and environmental sciences are all on the rise," explains **Kang Shichang**, director of the CAS State Key Laboratory of Cryospheric Science in Lanzhou. In some cases, he adds, opportunities and grants for research here are more generous than in the United States. He predicts that as Chinese funding for research balloons, more of his compatriots at American universities are likely to follow in his footsteps by returning to China.

Two million yuan (US$323,000) in research funds covering a four-year period, plus 700,000 yuan (US$114,000) in startup seed money, impelled Kang continued>

中国科学院大连化学物理研究所
Dalian In statute of Chemical Physics, Chinese Academy of Sciences





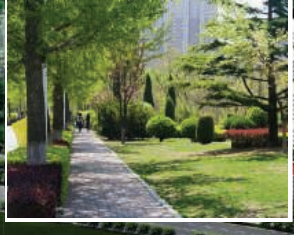












Welcome Talents to Joining Dalian Institute of Chemical Physics, Chinese Academy of Sciences

Dalian Institute of Chemical Physics

Founded in 1949, Dalian Institute of Chemical Physics (DICP), Chinese Academy of Sciences (CAS), is located in the beautiful port city of Dalian in Northeast china. In the past half century, research here has closely reflected the economic and scientific needs of China. The Institute has an important series of achievements, principally in the fields of catalytic chemistry, chemical engineering, chemical laser and molecular reaction dynamics, organic synthesis, and chromatography for modern analytic chemistry and biotechnology. These achievements have contributed greatly to the economic and technological development of China.

Dalian National Laboratory for Clean Energy

Dalian National Laboratory for Clean Energy (DNL)is China's first National Laboratory in the field of energy research, and integrates laboratories across DICP and other institutions. Research at DNL is focused on the efficient conversion and optimal utilization of fossil energy, clean energy conversion technologies and the economically viable use of solar and biomass energy. The research activities will benefit from the existing strong background of the DICP in chemistry, physics, chemical engineering, biology, and materials science, in addition to our extensive network of national and international collaborations. DNL is integrated on the DICP campus and would be taken as a main research platform to develop both fundamental and applied researches in DICP.

Development Strategy

DICP focuses on sustainable energy research and coordinates the development of environment optimization, biotechnology and advanced material in a multidisciplinary atmosphere by strengthening technological integration and innovation. DICP aims to play indispensable roles in national economy and security, and to become a leading research institute in the world.

Research Fields

Physical Chemistry, Catalytic Chemistry, Industrial Catalysis, Petrochemical Engineering, Solar Energy, Photoelectrocatalysis, Biomass Conversion, Low-carbon Chemistry, Fuel Cell Technology, Energy Storage Technology, Materials Chemistry, Advanced Materials, Energy Materials, Organic Chemistry, Biotechnology, Bioinformatics, Structural Biology, Synthetic Biology, Translational Medicine, Metabonomics, Systems Biology, Cell Biology, Analytical Chemistry, Combustion Chemistry, Energy Strategy

Recruitment Programs

National Thousand Talents Program (Full-time & Part-time)
National Thousand Talents Program for Foreign Experts (Full-time)
National Thousand Talents Program for Young Scientists (Full-time)
CAS Hundred Talents Program (Full-time)
DICP Hundred Talents Program (Full-time)

Support, salary and benefits

Generous start-up package, competitive salary and benefits will be provided based on different recruitment programs. The start-up package includes adequate start-up funds, newly renovated office, ample laboratory space, and plenty of research assistants. The benefits include medical insurance, retirement and house subsidy. For leading professors, salary at a level of international counterparts would be offered.

Contact us

Please send applications, inquires and nominations to talents@dicp.ac.cn. All applications should include a CV and a list of detailed publications. Those interested in cluster hiring should send CVs and publications as a group. Visit DICP website for more details:http://www.talent.dicp.ac.cn.

DICP Outstanding Postdoctoral Scholarship

The "DICP Outstanding Postdoctoral Scholarship" provides RMB 1,000,000 to support 10 outstanding postdoctoral applicants with RMB 100,000 per year for each selected applicant during his/her postdoctoral period of two years. The salary for an outstanding postdoctoral fellow is in the range of RMB 180,000~200,000 per year.

DICP has two postdoctoral stations: Chemistry and Chemical Engineering & Technology. DICP's postdoctoral program includes research in all areas represented at DICP.

To apply for a postdoctoral position, send a cover letter, CV, and the names and contact information for three references to Dr. Hao Yu （postdoc@dicp.ac.cn）

ScienceCareers online @sciencecareers.org
FOCUS ON CHINA

The mountains that surround Lanzhou and the deserts that cut through Gansu Province create an oasis for academic study that contrasts with the frenetic cities and colleges along China's cosmopolitan coast, explains Hu Bin.



to relocate from a postdoctoral research position at the University of Maine to set up a lab focusing on climate change across the glaciers, lakes, and mountains that cap the Tibetan Plateau. Now, he leads frequent expeditions to the highest peak on the planet—Mount Everest—to carve out ice cores that carry frozen traces of modern pollutants and glaciochemical records of centuries past.

His lab is part of Lanzhou's Cold and Arid Regions Environmental and Engineering Research Institute, which hosts more than 40 foreign experts who either serve on an advisory committee or as guest professors.

The region's astounding topography also holds appeal for researchers beyond the environmental sciences. The mountains that surround Lanzhou and the deserts that cut through Gansu Province create an oasis for academic study that contrasts with the frenetic cities and colleges along China's cosmopolitan coast, explains **Hu Bin**, dean of the Lanzhou University's School of Information Science and Engineering. Enrollment is exploding at the school, as well as in biomedical engineering and telecommunications, he says.

Previously a visiting scholar at the Swiss Federal Institute of Technology in Zurich, Hu accepted his position as part of the central government's Thousand Talents Program and will continue his research on Semantic Web projects, which range from digital publishing to Web-augmented diagnosis tools for hospitals.

In another sector of the city, Lanzhou's Institute of Modern Physics (IMP) is offering packages through the CAS Hundred Talents Plan to attract outstanding experts in hadron physics and high-profile accelerator physicists.

IMP Director **Xiao Guoqing**, who obtained a doctorate in physics from the Free University of Berlin, explains the institute set up a special department to focus on research and development of a superconducting proton collider.

Leading physicists across the country and around the world have been tracking China's plans to construct a Circular Electron Positron Collider as well as a Super Proton Proton Collider inside the same underground tunnel, adds Xiao.

It would be "very, very good" for the province if the supercollider could be located in Gansu, he says.

Creating this supercollider, measuring up to 80 km in circumference, "would undoubtedly attract many particle physics and (most importantly) accelerator physicists to help in the construction of the accelerator and the mounting of experiments," predicts **David Gross**, a Nobel laureate in physics and scholar at the Kavli Institute for Theoretical Physics, at the University of California, Santa Barbara.

"Physicists go to where the best physics can be done," says Gross, who traveled to China earlier this year to join a symposium on the twin colliders.

Designing and constructing China's circular colliders will rely on an international team of scientists coalescing around one of the most ambitious projects centered on peace and science in human history, explains **Gao Jie,** one of the primary shapers of the future accelerators at the CAS Institute of High Energy Physics.

"We will welcome scientists across the world to participate," says Gao, who conducted research as an accelerator physicist at the Linear Accelerator Laboratory, National Institute of Nuclear and Particle Physics (IN2P3), in the French city of Orsay for more than 15 years.

In one of the first steps toward constructing these colliders, Professor Gao's institute has created the Center for Future High Energy Physics in Beijing, headed by **Nima Arkani-Hamed**, one of the foremost theoretical physicists in the United States.

Arkani-Hamed, a scholar at Princeton's Institute for Advanced Study, says the new ringed accelerators could attract up to 10,000 scientists from across the planet, and will make China "the collider center of the world."

Nobel Laureate Gross agrees: "There are few areas where a single country can leap into the lead role in a fundamental area of research—this is one."

Taking the lead in the future might depend in part on breathing new life into the past. Ancient Silk Road connections with other cultures helped the western sector of the Chinese empire create some of the most advanced cities and learning centers in human history. The capital of the Tang Dynasty (618–907), in western China, was by many measures the most cosmopolitan city of its time, with the Silk Road attracting not just traders, but scholars and philosophers, into the region. Western China's present-day efforts to create a "New Silk Road" that links up all Eurasia and to attract scholars from across China and around the world may similarly transform the region into an expanding oasis for academia and science.

Kevin Holden is a science writer based in Beijing, China.

DOI: 10.1126/science.opms.r1400151




The auditorium, erected in 1931


The Li Wenzheng Library

Recruitment of Global Talents, Southeast University

Established in 1902, located in Nanjing, formerly known as Nanjing Institute of Technology in 1952, Southeast University is one of the top research-oriented universities in China, focusing on fundamental cutting edge research and high-level education, covering such diverse fields as science, engineering, technology, humanities, economics, management and law. SEU is one of the first universities funded by China's "211" and "985" programs. It has over 20 national or provincial research institutes, with a number of key research bases. (www.seu.edu.cn)

Requirements & Qualifications

Southeast University invites applications from outstanding scientists for tenure track positions, and positions of chair professors and visiting professors. The University also provides distinguished candidates with very strong continuing support to apply for programs such as "1000 Plan Professorship for Young Talent" and "Chang Jiang Scholars Program".

1000 Plan Professorship for Young Talent

Candidates should be under the age of 40; in engineering or natural sciences major; with Ph.D. or M.D./Ph.D. degree from a world-renowned university with at least 3 years of research experience, or with PhD degree from Mainland China with at least 5 years of research and teaching experience abroad. Applicants should have a position in a well-known university or other corresponding institutions overseas.

Chang Jiang Scholars Program(Distinguished Professors/Chair Professors)

Distinguished Professors: Applicants should hold at least associate professor or other equivalent positions in well-known universities. Besides, applicants should be below 45 years old if in a field of natural sciences, and below 55 in a field of humanities and social sciences. Successful applicants should work full-time in the university. The employment contract is for 3 years.

Chair Professors: Applicants should have a strong research background as well as teaching experience, and should currently hold professorship or other corresponding positions in well-known universities abroad. Successful applicants should work in the university for at least 2 months per year. The employment contract is for 3 years.

Salary, Benefits and Support

Successful applicant will be offered highly competitive salary and benefits, research space, extensive opportunities for collaboration both within the University and with neighboring institutions. The support includes sufficient laboratory space, startup funding, relocation fee and competitive salary commensurate with experience, and other employee benefits.

Contact Us

The application period is not limited.

Qualified applicants for "1000 Plan Professorship for Young Talent" and "Chang Jiang Scholars Program" are invited to send the job intention letter and CV by email to the following contact person.

Equal Opportunity Employer for Minorities/Females/Persons with Disabilities/Veterans

Address:Human Resource Department, Southeast University, 2 Sipailou Rd, Nanjing, China , 210096
Contact Person:LIU Lili,SUN Yuman Tel:+86-25-8793301,+86-25-52090253
Fax:+86-25-83793301 Website:http://rsc.seu.edu.cn/s/98/t/81/p/2/c/1146/list.htm
E-mail:ymsun@seu.edu.cn

电子科技大学 Senior Faculty Openings at UESTC

University of Electronic Science and Technology of China

University of Electronic Science and Technology of China (UESTC) is a multidisciplinary research university included in "Project 211" and "Project 985", situated in Chengdu, Sichuan Province. It is founded in 1956 and renowned as China's cradle for the national electronic industry. The university boasts top one discipline of Electronic Science and Technology in China (China Discipline Ranking), while also having departments of engineering, management and medicine.

We aim to recruit candidates for China`s national talent plans. As a key university, we offer competitive salary, housing benefit and startup research funding package to qualified applicants:

◆Long-term Thousand Talent Plan: Ph.D. from overseas educational institutions, Professor or equivalent position, able to work fulltime during the tenure.

◆Thousand Young Talent Plan: Ph.D. from overseas educational institutions, at least 3-year overseas working experience (or Ph.D. from institutions in China with at least 5-year working experience abroad).

◆Foreign Experts of Thousand Talent Plan: Non-Chinese experts, Professor or equivalent position in prestigious overseas universities or R&D institutes.

◆Accredited Professors of Cheung Kong Scholar's Program: Professor or equivalent position if working in China, able to work fulltime during the tenure (associate professor or equivalent position if working overseas).

◆Chair Professors of Cheung Kong Scholar's Program: Professor or equivalent position in top overseas institutes, able to work for more than 2 months in China annually.

◆Hundred Talent Plan of UESTC: Ph.D. from overseas educational institutions, potential candidate for national talent plans of China.

There are also faculty openings for:

◆Professor: Ph.D. from overseas educational institutions, professor title or the equivalent in prestigious institutes, excellent academic record.

◆Associate Professor: Ph.D. from overseas educational institutions, proven academic record, and great potential in achievement.

◆Assistant Professor: Ph.D. from overseas educational institutions, great potential of success in academia.



For more info, please contact us at: uestc-hr@hotmail.com OR zhaopin@uestc.edu.cn
Official website: http://www.oice.uestc.edu.cn/en/ OR http://www.jobs.uestc.edu.cn/

Teaching and Research Positions at Xuzhou Medical College









Xuzhou Medical College, situated in the city of Xuzhou, was founded in 1958.It is now the center of medical education, medical service and medical research in Northern Jiangsu and in the whole Huaihai economic zone.

XZMC offers 20 undergraduate programs, such as Clinical Medicine, Anesthesiology, Medical Imageology, Pharmacology, Stomatology, Nursing, Biomedical Engineering, Public Administration, etc., 2 first-level discipline doctoral programs, namely Clinical Medicine and Biology and 5 first-level discipline master's programs, namely Clinical Medicine, preclinical medicine, Biology, Pharmacology, public health and preventive medicine. Still, the college has post-doctoral research centre.

I.Disciplines for Application:

Biology, Basic Medicine, Pharmacy, Clinical Medicine,Public Health and Preventive Medicine, Stomatology,etc..

II.Requirements for overseas applicants:

1)Doctoral degree holder
2)Professor or associate professor

III. More information and to apply visit website:

http://rsc.xzmc.edu.cn/

IV. Contact:

Mr. Du Gang or Mr. Li Yuandong
Personnel Department, Xuzhou Medical College
209 Tongshan Road, Xuzhou, Jiangsu Province, 221004, China
Tel/Fax:+86-516-83262042, 83262238
Email: szb@xzmc.edu.cn rsc@xzmc.edu.cn

Southwest Jiaotong University, P.R.China Anticipates Your Working Application







Southwest Jiaotong University (SWJTU), founded in 1896, situates itself in Chengdu, the provincial capital of Sichuan. It is a national key multidisciplinary "211" and "985 Feature" Projects university directly under the jurisdiction of the Ministry of Education, featuring engineering and a comprehensive range of study programs and research disciplines spreading across more than 20 faculties and institutes/centers. Boasting a complete Bachelor-Master-Doctor education system with more than 2,500 members of academic staff, our school also owns 2 first-level national key disciplines, 2 supplementary first-level national key disciplines (in their establishment), 15 first-level doctoral programs, 43 first-level master programs, 75 key undergraduate programs, 10 post-doctoral stations and more than 40 key laboratories at national and provincial levels.

Our university is currently implementing the strategy of "developing and strengthening the university by introducing and cultivating talents". Therefore, we sincerely look forward to your working application.

More information available at http://www.swjtu.edu.cn/

A.High-level Leading Talents

It is required that candidates be listed in national top talents programs such as **Program of Global Experts, Top Talents of National Special Support Program, "Chang Jiang Scholars", China National Funds for Distinguished Young Scientists and National Award for Distinguished Teacher.**

Candidates are supposed to be no more than 50 years old. The limitation could be extended in the most-needed areas of disciplinary development.

Candidates who work in high-level universities/institutes and reach the above requirements are supposed to be no more than 45 years old.

B. Young Leading Scholars

Candidates are supposed to be listed in or qualified to apply for the following programs:

•National Thousand Young Talents Program

•The Top Young Talents of National Special Support Program
(Program for Supporting Top Young Talents)

•Science Foundation for the Excellent Youth Scholars

Candidates should have good team spirit and leadership, outstanding academic achievements, broad academic vision and international cooperation experience and have the potential of being a leading academic researcher.

C. Excellent Young Academic Backbones

Candidates under 40 years old are expected to graduate from high-level universities/institutes either in China or other countries. Those who are professors, associate professors and other equal talents from high-level universities/institutes overseas could be employed as professors and associate professors as well.

D. Excellent Doctors and Post Doctoral Fellows

Candidates under 35 years old are supposed to be excellent academic researchers from high-level universities either in China or other countries.

The candidates will be provided with competitive salaries and welfares that include settling-in allowance, subsidy of rental residence, start-up funds of scientific research, assistance in establishing scientific platform and research group as well as international-level training and promotion . As for outstanding returnees, we can offer further or specific treatments that can be discussed personally.

Contacts:

Ye ZENG & Yinchuan LI
Telephone number: 86-28-66366202
Email: talent@swjtu.edu.cn
Address: Human Resources Department of SWJTU, the western park of high-tech zone, Chengdu, Sichuan, P.R.China, 611756

河海大学 HOHAI UNIVERSITY







Hohai University, a centennial famous university, is one of the national key universities administered directly under the Central Government and the Ministry of Education of China. It has become a comprehensive and research-oriented university featuring the coordinated development of multi-disciplines, with Hydrology as its characteristics and engineering as focus. It is one of the universities of Project 211, 985 Project Innovation Platform. Also it is one of 56 universities approved on theestablishment of graduate school.

As the earliest and largest university engaged in researching and teaching hydraulic engineering and other related fields, Hohai university now comprises national key laboratories of hydrology and water resource and engineering science, and national engineering research center of effective utilization of water resources and engineering safety. Also, Hydraulic Engineering with an international reputation ranks first in the discipline rankings evaluated by the academic degree and graduate education development center of Ministry of Education of China. Otherwise, the comprehensive strength of civil engineering, environmental science and engineering, mechanics, industrial and commercial management disciplines are among the top ones in china.

Meanwhile, Hohai has other relevant supporting disciplines with strong strength, such as, marine science, geological resources and geological engineering, surveying and mapping science and technology, electrical engineering, software engineering, computer science and technology, electronic science and technology, power engineering and engineering thermal physics, control science and engineering, transportation engineering, agricultural engineering, mechanical engineering, management science and engineering, new energy, new materials, mathematics, statistics, sociology, Marx theory, public management, applied economics, law, education and so on.

To further implement the strategy of "strong university with talents", to promote the construction of discipline and high level of qualified personnel, to construct an excellent and innovate University with aim of "first-class at home, world-renowned at abroad", Hohai University is perennially recruiting talents from overseas well-known universities and scientific research institutions.

1. Once the applicant of 1000 Talents Program (including Young Overseas High-Level Talents Introduction Plan), The Chang Jiang Scholar Program (distinguished professors and visiting professors) and other talent programs are employed, Hohai University will provide research funds, expenses for research starting and platform construction, offices, laboratories, equipment, and research assistants. Hohai will also provide support in terms of housing (Hohai provides apartment or implements housing monetized policy), sponsor's job, entrance of children and etc. For those who are extraordinarily excellent, their remuneration would be correspondingly adjusted.The positions of 1000 Talents Program, The Chang Jiang Scholar Program and other programs are open for you. Please contact us:

Contact persons: Ms. Qiao, Mr. Zhang
Main Email: 443979815@qq.com, lingbinzhang@qq.com Tel: 86-25-83786484

2. Excellent doctors, post-doctors, full-time teachers and researchers who want to candidate positions of teaching, please visit the website: http://rsc.hhu.edu.cn/ for more detailed information of the 2015 recruitment plan and contact principals of responding college.

Based on the characteristics of the academic disciplines and actual needs, the applicant, when employed, will be provide with including salaries, research starting funds, fees for renting or buying an apartment and so on. Otherwise, the excellent will be provided with training plans such as Young Professors, Innovation Plan, Leading Talents, and Hohai Scholars.

Address: Department of Personnel, Hohai University, No.1 Xikang Road, Nanjing, Jiangsu Providence, China
Postal Code: 210098
Contact persons: Ms. Xu, Mr. Wu
Tel: 86-25-83786205
Main Email: wgzhhu@hhu.edu.cn, pretty_weiwei@126.com Fax: 86-25-83786767
Please send e-mails or call us to get more information about requirements and remuneration.

中國藥科大學

Faculty Positions Available in China Pharmaceutical University











China Pharmaceutical University (CPU), founded in 1936, is one of the "211 Project" key universities with top-ranking pharmaceutical related disciplines and innovative platforms affiliated to the Ministry of Education of China. Located in Nanjing, Jiangsu Province, CPU was China's first independent school of pharmacy and currently consists of 10 pharmaceutical related schools and departments. With spissated atmosphere in pharmaceutical science, CPU laid its leading position in China with consecutively national 1st prize in pharmaceutical education. More than 10 national, provincial or ministerial key laboratories guarantee researchers to receive and fulfill various key funding projects in innovative drug developments. The recent established Engineering School will replenish university's pharmaceutical technique chain to expedite pharmaceutical and medical development.

CPU boasts a highly qualified and experienced team of faculty and advanced research platform. To facilitate our further development, CPU focuses on talent team expanding. With our favorable learning and working environment, we invite outstanding scholars of all nationalities to join us. For more information about CPU, please visit http://www.cpu.edu.cn/.

CPU is seeking excellent talents in:

1) Pharmaceutical Science (including chemical drug development, Traditional Chinese Pharmacy, Biopharmaceutics)
2) Some specific area (Stem Cell, Omics Studies, Antibiotic Resistance, immunology, bioinformatics, molecular biology, chemical biology)
3) Pharmaceutical oriented engineering, including Biomedical Engineering (Molecular Imaging and nanomedicine, Molecular Biology, Computer assisted Drug Design, pharmacology), Pharmaceutical Engineering, Environmental Engineering, Food and drug quality and evaluation.

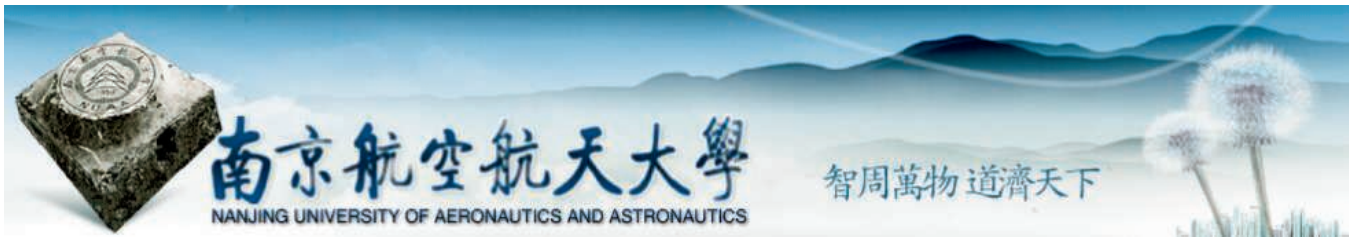
Minimum Qualifications:

Qualified candidates must have an earned Ph.D. degree, postdoctoral training experience or work experience in related research areas. Applicants for Professors should have an established research program with funding history, independent research and teaching experience in a major research institution with an outstanding publication record.

Contact Us:

Qualified applicants are invited to email their CV, which should include the basic information, education history, publication record, interested position, and statement of research interests to rcb@cpu.edu.cn. Applications will be reviewed on a rolling basis until the positions are filled.
Phone Number: +86 025 86185091










NUAA Hunting for Talents from Home and Abroad

Nanjing University of Aeronautics and Astronautics(NUAA), established in 1952, enjoys a well-balanced development of multiple disciplines in Natural Sciences, Humanities, Social Sciences, Management, etc. with the characteristics of Aeronautics, Astronautics and Civil Aviation. NUAA is among the first batch of national key universities entitled to award doctoral degree, master's degree. It is one of the 56 universities with graduate school in China, and it has also been listed as one of the "National Project 211" universities.

NUAA will provide teachers and researchers with a good academic environment, satisfactory working and living conditions and a stage on which they can put their talents to good use.

1. Positions and Offer

1.1 NUAA hunts for high-level talents, including:

A. members of “Recruitment Program of Global Experts”, distinguished professors of “Chang Jiang Scholars Program”, gainers of “National Science Fund for Distinguished Young Scholars” and outstanding teachers at the national level.

The age is normally under 50, but the limitation of age can be moderately adjusted for a few disciplines which are in urgent need of talents.

B. some talents with great competition potential and eligibility for the 4 programs above, and will be employed as distinguished professors.

The age is normally under 42.

1.2 NUAA hunts for young talents, including:

Outstanding young scholars from home and abroad, to be trained to enter the “Chang Jiang Scholars Program”, gain “National Science Fund for Distinguished Young Scholars”, and assessed as a national famous teacher, etc.

Conditions of employment:

1.2.1. The talent should have rigorous style of study, and have both ability and political integrity with strong innovation and dedication.

1.2.2 He or she should have a broad academic vision, have achieved outstanding academic performance in the field engaged, and achieved academia peer’s recognition, with potential to be a leader in this field.

1.2.3 The talent should have innovative ideas in the development of this discipline and academic research, and having the ability to lead this discipline to catch up with or maintain the international advanced level.

1.2.4 The talent should have a doctorate, and be under the age of 38.Please find details about the offers in Table 1.1

Table 1.1

Category	Pay (gross)	Housing	Fund for Starting Work	Other
High-level talents A	Approx. 1.5 million yuan in total for an employment term of 3 years	An apartment (about 160m² with the market value of 2 million yuan) and 300,000 yuan for settling	Working conditions, fund for starting work, expenditure for discipline construction, team construction and graduate recruitment vary according to personal conditions.	NUAA provides medical treatment according to national regulations and green channel in campus hospitals; helps the spouse to find a job and the education of children, etc.
High-level talents B	Approx. 1 million yuan in total for an employment term of 3 years	An allowance of 1 million yuan for housing; the employee can also apply to buy an apartment built by NUAA with a discount based on the market value		NUAA provides medical treatment according to national regulations and green channel in campus hospitals; helps to deal with the education of children, etc.
Young talents	Approx. 0.7 million yuan in total for an employment term of 3 years	An allowance of 0.7 million yuan for housing		NUAA provides medical treatment according to national regulations and green channel in campus hospitals; helps to deal with the education of children, etc.

1.3 NUAA hunts for teachers, including:

A. doctors from Chinese or foreign high-level universities or excellent masters (only for a few disciplines);

B. doctors with over 2 years of teaching and research experience in Chinese or foreign high-level universities or famous research institutions;

C. lecturers, associate professors and professors (or equivalent professional titles) from Chinese or foreign high-level universities or famous research institutions, or people who are excellent in teaching or research.

Please find details about the offers in Table 1.2

Table 1.2

Category	Fund for Starting Work		Housing Allowance (unit:1,000 yuan)	Other
	Engineering (unit:1,000yuan)	Other Disciplines (unit:1,000 yuan)		
Doctors/Lecturers	50-60	30-40	320	NUAA provides temporary apartments or single apartment renting
Associate professors	60-80	40-50	370	
Professors	100-200	50-100	470	

2. How to apply

2.1 For those who apply for posts in Table 1.1, please send resume (including basic information, education background, main papers and works, research and awards, etc) to zhaohaiyan@nuaa.edu.cn. Please clarify the post in the email. And the email will be replied within 3 work days.

2.2 For those who apply for the posts in Table 1.2, please log onto the website http://zp.nuaa.edu.cn(website of talent recruitment of NUAA). Please fill in personal information and the “application form for teaching post of Nanjing University of Aeronautics and Astronautics”.

2.3 Posts above are available throughout the year.

3. Contacts

Ms. ZHAO Haiyan, Mr. CAO Yunxing
Personnel Division, NUAA
Address: 29# Yudao St. Nanjing, Jiangsu Province, Postcode: 210016
Tel: +86-25-84892461
Fax: +86-25-84895923
Email: zhaohaiyan@nuaa.edu.cn
Web:http://rsc.nuaa.edu.cn




哈爾濱工業大學

Faculty Positions Available at Harbin Institute of Technology: Harbin, China



Harbin Institute of Technology(HIT), subordinated to the Ministry of Industry and Information Technology(MIIT), was founded in 1920 and is mutually constructed by MIIT, Ministry of Education(MOE) and Heilongjiang Province. HIT is one of the first universities supported by the "211 Project", "985 Project" and "2011 Project" of Chinese government. Over more than 90 years, HIT has developed into a distinctive, multi-disciplinary, open, research-based national key university. Our goal is to become a world-class university. You are welcomed to join us to realize your dream.

Requirements and Responsibilities

HIT invites applications worldwide for full-time faculties. Responsibilities include participating in teaching at the undergraduate and graduate levels and establishing high-level research programs. Candidates with a PhD degree may apply suitable positions mainly in the disciplines below:

Aeronautical and Astronautical Science and Technology/Mechanics
Mechanical Engineering/Automatic Test and Control
dynamics & control/Materials Science and Engineering
Power Engineering and Engineering Thermal Physics/Computer Science and Technology
Civil Engineering/Management Science and Engineering
Chemistry/ Physics/ Mathematics
Life Science and Technology

All positions include three levels--lecturer, associate professor and professor. Especially, exceptional candidates who meet relevant criteria will be recommended and encouraged to apply for competitive talent programs as follows:

A. 1000 Plan Professorship for Young Talent

Eligible candidates should be under the age of 40 and work in any field of natural sciences or engineering technology, with a PhD degree awarded in a world-renowned institution with at least 3 years of overseas research experience. Candidates should also be holding a full-time teaching or research position in renowned universities, research institutions, or R&D departments in famous overseas enterprises at the time of application.

B. "Outstanding Young Experts" Program Professor & Associate Professor

Eligible candidates should have a Ph.D. degree in world-famous universities and high-level achievements among peer researchers in China. Candidates will start a full-time work at HIT in less than half a year after a successful application.

Compensation and Benefits

General compensation and benefits are in line with current standards of HIT. Successful candidates who are inducted into "1000 Plan Professorship for Young Talent" or "Outstanding Young Experts" Programs can be provided with a treatment according to the programs including a start-up fund (including relocation and housing allowance) based on their research areas and academic experience.

Application Procedure

The posted positions are all immediately available until filled. Application package including a cover letter declaring the targeted school/research area and position, a detailed resume, a publication list, a research and teaching plan should be sent by email with the title of "Faculty Application from ScienceCareer". A prescreening decision reply will be provided as soon as possible upon the application is received.

To know more about the positions of HIT, please visit:
http://www.hit.edu.cn/express/employ.htm

Contact

Mr. Wang Yuebang or Lv Yueyong（Tel.:+86-451-86418762）
Mr. Wang Liang （Tel.:+86-451-86418579）
Fax.:+86-451-86414661
Email:jobs@hit.edu.cn

東華大學

Faculty Position Available in Donghua University

Donghua University, located in Shanghai， is one of the national key universities under the direct administration of the Ministry of Education since 1960. It is a member of the Project 211 universities. It was founded in 1951 as East China Textile College. In 1985, it changed its name to China Textile University, and to its present name, Donghua University in 1999. It is one of the first universities accredited by the Ministry of Education for granting the doctor, master and bachelor degrees.

Donghua University has growing into a distinctive multi-disciplinary university, with engineering as the predominant discipline and coordinated development of engineering, science, management and liberal arts disciplines in the past 60 years. The university has 5 post-doctoral research stations, 7 first-grade authorized doctoral programs, 24 first-grade authorized master programs, 6 categories of professional masters, 17 authorized master programs of engineering, and 56 majors for undergraduates in about 10 different disciplines. The university has one First-grade National Key Discipline, 5 Second-grade National Key Disciplines, 1 National Key (Cultivating) Discipline, 7 Shanghai Key Disciplines, 12 national, provincial or ministerial level scientific research bases, 2 bases of the Discipline Innovative Engineering Plan launched by the Ministry of Education and 1 National University Science Park. The university now has about 30,000 enrolled students, including about 15,000 undergraduates, 6,000 graduates, 5,000 continuing education diploma students and 4,000 international students. The university has now boast about 2,300 faculty and staff members. Among the 1,200 faculty members, there are 8 academicians of the Chinese Academy of Sciences or the Chinese Academy of Engineering and over 800 senior professors, including winners of the National Thousand-talent Project, the Changjiang Scholar Project and the National Science Fund of Distinguished Young Scholars.

Donghua University invites applications for full-time Professors， Associate Professors and excellent scientists. Successful candidates will be provided competitive salaries and start-up funds.

Positions Available :

Position offered by the Recruitment Program of Global Experts (1000 Plan Professorship)
Position offered by the Chang Jiang Scholars Program
Position offered by the Recruitment Program of Global Young Experts (1000 Plan Professorship for Young Talents)
Position offered by Donghua University Distinguished Research Fellow

Interested individual should send curriculum vitae by email to rcb@dhu.edu.cn. For more information, please visit the university Human Resource Department website http://web.dhu.edu.cn/rcbdhu/， or contact us by email rcb@dhu.edu.cn or by telephone 021-67792042.

哈爾濱工業大學
深圳研究生院

Dean and Heads for School of Science Wanted in Harbin Institute of Technology Shenzhen Graduate School





Approved by the Ministry of Education of China, Harbin Institute of Technology Shenzhen Graduate School was co-founded by the Shenzhen Municipal Government and Harbin Institute of Technology in Feb. 2002. It is an institution for full-time graduate education in 22 disciplines. There are 364 faculty members and staff as well as 2938 students on campus now. In May 2014, the school got approval from the Ministry of Education of China to be developed into Harbin Institute of Technology (Shenzhen) which will include undergraduate programs with natural sciences and engineering as its focus.

Dean for the School of Science, heads for departments of mathematics, physics, chemistry, and Center for Life and Health Medicine

Requirements

(1) Holder of doctoral degree; professor or at equivalent position with great influence and good reputation; priority be given to members of the national "Recruitment Program for Global Experts", winners of the "National Science Funds for Distinguished Young Scholars", and distinguished professors from the "Chang Jiang Scholars Program" of Ministry of Education of China;
(2) 3 or more than 3 years' management experience as department head or position above in renowned domestic or oversea universities;
(3) Full time working in the school; willing to undertake management responsibility for the school or the department;
(4) Good physical and mental condition; under the age of 50; applicants over 50 might also be considered if he/she is particularly excellent.

Employment & Benefits

(1) Way of Employment: An employment agreement will be signed where responsibilities, rights, and benefits of the position would be stated. The appointment will initially be made on a five-year fixed-term;
(2) Salary & Benefits: not less than 800 thousand RMB/year (negotiable), and relevant benefits such as housing subsidies and research funds for high-level talents according to policies of Harbin Institute of Technology and that of Shenzhen Municipal government.

More positions are available for:

*Members of the national "Recruitment Program for Global Experts"(1000 Plan Professorship)
*Winners of the "National Science Funds for Distinguished Young Scholars"
*Professors of the "Chang Jiang Scholars Program" of Ministry of Education of China
*Members of the national "Recruitment Program for Global Experts"(Youth)

Applicants who apply for dean for the School of Science or heads for departments of mathematics, physics, chemistry, and Center for Life and Health Medicine, please send the e-copy of your application materials and an Employment Work Plan to ztyszsz@hit.edu.cn or hr@hitsz.edu.cn before 31st Dec. 2014.Other positions are open all the year round. For more information, please visit: http://www.hitsz.edu.cn/body/rencaizhaopin/.

Contact

Mr. Tongyin Zhao
E-mail：ztyszsz@hit.edu.cn/hr@hitsz.edu.cn
Tel：+86-755-26033739/26033365
Fax：+86-755-26033800
Website：http://www.hitsz.edu.cn/

西北农林科技大学
NorthWest A&F University

NWAFU Professorship's Recruitment

Northwest A&F University (NWAFU) is a national key comprehensive university directly under the jurisdiction of the Ministry of Education of the People's Republic of China and supported by its "Project 985" and "Project 211". The University is located in Yangling, Shaanxi Province, 80 km west of Xi'an, a historic city of China. Yangling is not only the location of the University, is also the birth place of Chinese agricultural civilization and the location of the state-level Agricultural Hi-tech Industries Demonstration Zone. The University recruits distinguished professors and full-time professors in various academic disciplines.

I. Academic Disciplines:

Biology, Agricultural Engineering, Food Science and Engineering, Crop Science, Horticulture, Pedology, Plant Pathology, Entomology, Pesticide Science, Animal Husbandry, Veterinary, Machinery Design & Manufacture, Computer Science & Technology, Agricultural & Forestry Economic Management, etc.

II. Applicant's Eligibility

Candidates must possess a position at Assistant professor or higher ranking in a recognized college or university, and has achieved significant accomplishments or contributions, or evidence of success in the respective discipline.

III. Facilities and Conditions

The University provides sufficient start-up funds and facilities, highly competitive salary, and an apartment with individual property right after serving the University for 10 years. The University will actively help and recommend the hired professors to apply "Global Experts Program" and "Chang Jiang Scholars Program", and many others.

IV. Procedure

For full consideration, interested applicants should send a letter of interest, a recently updated CV with a list of publications, faculty position application form (http://rcb.nwsuaf.edu.cn/sort.php?sortid=8), and three letters of recommendations to: rencaike@nwsuaf.edu.cn. The University will arrange an interview after verifying of the materials and application materials. The University will cover the expenses of international travel when the candidate is invited for an interview.

Please visit http://rcb.nwsuaf.edu.cn/ for more information.

ScienceCareers online @sciencecareers.org
FOCUS ON CHINA

清華大學深圳研究生院
Graduate School at Shenzhen, Tsinghua University

ACADEMIC AND RESEARCH POSITIONS AT GRADUATE SCHOOL AT SHENZHEN, TSINGHUA UNIVERSITY

Graduate School at Shenzhen, Tsinghua University was jointly established by Tsinghua University and Shenzhen Municipal Government. Situated in Shenzhen University Town, Shenzhen, Guangdong Province, China, the School is setting up three multidisciplinary research centers in Nanotechnology and Nanomedicine, Low-Carbon and New Energy Technologies, and Information Science & Technology; and is now recruiting faculty and researchers globally to conduct innovative, cutting-edge researches.

Successful candidates must have a Ph.D. degree from an accredited university, a demonstrated ability for independent research and graduate teaching, and have achieved significant research outcomes. Candidates applying for PI positions should demonstrate their pioneering research outcomes, peer recognition from home and abroad, and capability to lead the research group to reach international advanced level.

The positions are open to the candidates in the following areas: Environmental Science and Engineering, Mechanical Engineering, Instrument Science and Technology, Material Science and Engineering, Electrical Engineering, Information and Telecommunications Engineering, Control Science and Engineering, Computer Science and Engineering, Physics, Chemistry, Biology, Management Science and Engineering, Logistics Engineering and Management, Sociology, etc.

The School will provide competitive salary ranging from CNY 400,000 to CNY 1,200,000; the School will also assist the faculty and researchers to apply for housing allowance up to CNY 1,500,000 sponsored by Shenzhen Peacock Project and Shenzhen High Level Talents Plan. The post-doctoral researchers will receive living allowance of CNY 120,000 per year paid by Shenzhen Government, plus salary paid by the School.

The School is also recruiting students globally for the Ph.D. programs, who will be enrolled in September, 2015 to conduct multidisciplinary study and research. Upon meeting program requirements and passing in the dissertation defense and review, students will be awarded Diploma and Ph.D. Degree by Tsinghua University.

To apply for the academic and research positions, please send the following documents to tucy@sz.tsinghua.edu.cn:
1.Curriculum vitae
2.Application form
3.3 to 5 representative publications
4.Contact information of 5 to 10 preferred international reviewers including mailing addresses, e-mail addresses, and phone numbers
5.Future work plan

For any enquiries please contact Ms. Tu via at tucy@sz.tsinghua.edu.cn or telephone +86-755-26036856.

To apply for the Ph.D. programs, please contact Ms. Yang at +86-755-26036113, or Ms. Zhao at +86-755-26036020, or send email to yanggd@sz.tsinghua.edu.cn.

For more information about the School, please refer to the official website: www.sz.tsinghua.edu



Faculty Recruitment- Professors / Associate Professors / Assistant Professors

School of Advanced Computing, Sun Yat-sen University



Sun Yat-sen University(SYSU) was founded in 1924 by Dr. Sun Yat-sen (also called Sun Zhongshan), a great democratic revolutionary leader of the 20th century. The University is located in Guangdong Province, an area neighboring Hong Kong and Macao, which is at the forefront of China's reform and opening up. Being one of the leading universities in the People's Republic of China, SYSU is a comprehensive multi-disciplinary university, including the humanities, social sciences, natural sciences, technical sciences, medical sciences, pharmacology, and management sciences. (For further information, please visit http://www.sysu.edu.cn)

The School of Advanced Computing (SAC) was founded in March 2013. The school associates with the "National Supercomputing Center in Guangzhou" and several national/ministerial and provincial-level research laboratories or institutes, and processes "Milkyway-2", the fastest supercomputer in the world. The school runs with the sole aim of conducting world-class academic research in computer science and cultivating a new generation of scientists with international recognition. In addition to prompting scientific research, the school endeavors to address the real challenges in industrial applications by technical transferring.（For further information, please visit http://www.sc.sysu.edu.cn/）
The School is seeking for outstanding applicants in the all fields of Computer Science.

Job Description

Faculty Positions: Professors / Associate Professors /Assistant Professors
Location: Higher Education Mega Center (HEMC), Guangzhou City, China
Duties
The appointees will be required to:
(a) teach at both undergraduate and postgraduate levels; supervise student projects and theses;
(b) conduct research that leads to publications in top-tier refereed journals and awards of research grants;
(c) provide leadership in research, programme/curriculum development and administration;
(d) engage in research collaborations both internally and internationally;
(e) participate in professional services to the academic community and in promotional activities;
(f) perform any other duties as assigned by the Dean of School or his delegates.

Qualifications

(a) a PhD in a relevant discipline, plus substantial years of experience in teaching, research and industry;
(b) excellent communication skills and the ability to use English and Chinese as the medium of instruction.

Candidates with less experience will be considered for appointment at the level of Associate Professor /Assistant Professor.

Salary, Benefits and Support

A highly competitive remuneration package will be offered.

Contact Us

Applicants should send a detailed CV, teaching and research statements, and list of publications to Mr. Li (schr@mail.sysu.edu.cn). Confidential review of applications will begin immediately and until the positions are filled.

For further information, please visit http://www.sysu.edu.cn
TO APPLY, please contact:
Mr. Li, School of Advanced Computing at Sun Yat-sen University
E-mail: schr@mail.sysu.edu.cn
Telephone: +8620 39943563



陕西师範大學
SHAANXI NORMAL UNIVERSITY

Professorships and Chair Professorships at Shaanxi Normal University, China

陕西师范大学诚聘优秀人才







Founded in 1944 and located in the world-famous historical city – Xi'an, Shaanxi Normal University （SNNU) is one of the key institutions of higher learning directly affiliated to Ministry of Education and "211 Project University" in China. The university is seeking individuals with outstanding scientific credentials for The Recruitment Program of Global Experts(1000 talent plan), Chang Jiang Scholars Program, 100 Talent Program of Shaanxi Province and Qu Jiang Scholars Program of Shaanxi Normal University, which are designed for the recruitment at the level of professors, associate professors and chair professors, etc.

SNNU has a broad range of academic disciplines, and positions are available in all the relevant areas as below: Biology, Chemistry, Commercial Science, Communication, Computer Science and Technology, Economics, Education Science, Mathematics, Environmental Science and Engineering, History, Management, Material Science, Physics, Psychology, Western Literature, etc.

Qualifications

Applicants are expected to have remarkable academic achievements and to demonstrate capacity in leading an academic team to keep a competitive edge in frontier areas. Successful candidates for the professorship will be expected to undertake full-time teaching and research in general, and those for the chair professorship to work part-time (at least 2 months per year).

You are welcome to search for more information about the qualifications of each talent plan at http://rsc.snnu.edu.cn/browse.asp?id=2357

Salary and Housing Allowances

The university provides state-of-the-art research facilities and strong supporting staffs. Internationally competitive start-up support, salary and benefits will be offered according to qualifications and experience. Successful candidates of the specially listed programs will receive supplementary remuneration, including newly renovated office and laboratory spaces, and a highly collegial and interactive environment, as well as assistance on the establishment of a delicate research team.

Application Documents

Applicants are expected to submit a CV with cover letter which is supported by such documents as photocopies of advanced degrees, and three recommendation letters among other things.

Please direct your applications and inquiries to:
Mrs. Wu Jinfeng Miss Gao Kaixuan
Email address: rcb@snnu.edu.cn
Tel: 86-29-85310456, 86-29-85310455
Fax: 86-29-85310456

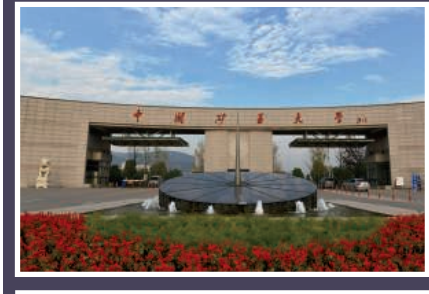


Recruitment of Global High-caliber Talents for CUMT











Located in the historically and culturally renowned city of Xuzhou, Jiangsu Province, CUMT is known as one of the national universities of 211 Project and 985 Project. The university has been fully equipped with advanced facilities which provide a favorable support for teaching and research. It now has 22 schools and 9 national key disciplines.

Requirements & Qualifications

CUMT is now recruiting both domestic and foreign professors, associated professors, assistant professors and outstanding young teachers with doctor degree, which provides a strong continuing support for applicants of "Thousand Talents Plan" and "Chang Jiang scholars Program".

Thousand Talent Program for Young Outstanding Scientists

Candidates should be under the age of 40; major in engineering or natural sciences with Ph.D. degree from a world-renowned university and with at least 3 years research experience. Applicants should have a position in well-known universities or other corresponding institutions overseas.

Chang Jiang Scholars Program (Distinguished Professors/Chair Professors)

Distinguished Professors: Applicants should hold at least associate professor or other equivalent positions in well-known universities. Besides, applicants should be below 45 years old in the field of natural sciences, and below 55 in the field of humanities and social sciences. Successful applicants would work at least 3 years.

Chair Professors: Applicants should have a strong research background as well as teaching experience, and should currently hold professorship or other corresponding positions in well-known universities abroad. Successful applicants should work in the university for at least 2 months per year. The employment contract is for 3 years.

Recruitment for Distinguished Professors of Jiangsu Province

Subjects: Mining and Underground Engineering, Resource Environment and Chemical Engineering, Mechatronical Science and Information Technology, New Energy Science and Technology,etc.

Qualifications: Applicants should below 45 years old, and hold assistant professorship or other corresponding positions in foreign high-level universities or scientific research institutions. Meanwhile, applicants should have a strong scientific research background as well as working experience abroad. Successful applicants should work full-time in the university. The employment contract is for 3 years.

Recruitment for Young Teachers

The university is recruiting outstanding Ph.D. graduates from first-class universities. The applicant is required to own abundant academic results and possess development potential in scientific research with several papers published in international well-known journals in recent years. Meanwhile, the applicant should be good at team work and has the experience of hosting or participating in key scientific research projects. For those applicants, please refer to the website:
http://pd.cumt.edu.cn/3e/68/c352a81512/page.htm.

Benefits and Salary

Successful applicant will be offered with highly competitive salary and benefits, research space, extensive opportunities for collaboration with CUMT and other neighboring institutions. Ph.D. graduates from famous universities abroad will be offered with competitive startup founding and relocation fee.

Contact us

The application period is not limited.

Qualified applicants for "Thousand Talent Program for Young Outstanding Scientists", "Chang Jiang Scholars Program", "Distinguished Professors of Jiangsu Province" and outstanding Ph.D. graduates, please e-mail your job intention letter and CV to the relevant college according to your major and the Division of Personnel separately (About the E-mail address of relevant college, please refer to recruitment website).

Address: Division of Personnel, CUMT, NO.1, Daxue Road, Xuzhou, China
Postcode: 211116
Contact Person: Shao Bo, Wang Pengpeng, Gao Ding
Tel: 0086-516-83590211
Fax: 0086-516-83590211
E-mail (the Division of Personnel): kdrc@cumt.edu.cn
Recruitment website: http://pd.cumt.edu.cn/
School website: http://www.cumt.edu.cn/

宁波大学
NINGBO UNIVERSITY







Ningbo Univerity invites you to apply for faculty positions

About Ningbo University

Founded in 1986 by Sir Yue-Kong Pao and autographed the name by Deng Xiaoping, Ningbo University (NBU) is a young and dynamic university located in the beautiful city of Ningbo by East China Sea, with five campuses covering 160 hectares of land. As a leading comprehensive university in Zhejiang Province, NBU offers programs in economics, law, education, liberal arts, history, science, engineering, agriculture, medicine, and management. The university now receives public funding, as well as continuous support and generous donations from many overseas Chinese and their families including Sir Yue-Kong Pao, Sir Run-Run Shaw, Chao An Chung, Hans Tang, Yue-shu Pao, Cao Guangbiao, Li Dashan, Zhu Xiushan, etc.

Academic Excellence

NBU consists of 22 faculties and offers 75 undergraduate programs, 116 master programs, and 12 Ph.D. programs. It enrolls 31,645 students including 26,527 full-time undergraduates and 5,118 graduate students. Currently NBU has around 1,400 full-time faculty members and 1,000 administrative staff members. Among them there are 5 academicians, 284 full professors, and 721 associate professors.

Research Achievements

With 77 research institutes and 14 key laboratories, NBU is the center of varieties of research and teaching activities. The research and development initiatives of the university, especially in marine science, information science and technology, engineering mechanics, and material science have contributed greatly to the economic development of the region and have been recognized by numerous national awards. The university library has a CNKI Network Administrative Service Center, and a collection of approximately 1,700,000 books and 11500GB digital resources.

International Programs

NBU maintains close links to 47 well-known institutions of higher education in Canada, Germany, France, Great Britain, USA, Sweden, Japan, South Korea and Australia. For example, the Sino-Canada joint-educational program is welcomed by international students with 100% satisfaction.

QUALIFICATIONS

Candidates should at least have (a) a Ph.D. degree in a related discipline, (b) adequate teaching ability and a strong passion for teaching, (c) an outstanding research background and influential publication record in recent three years, and (d) an ability to conducthigh-quality research and attract external funding.

REMUNERATION & CONDITIONS OF SERVICE

Salary offered will be commensurate with qualifications and experience. Remuneration package will be highly competitive. Forapplicants with titles of professor or associate professor, salary and housing compensation can be negotiated on the individual basis.For newly graduate PhD and Postdoctoral, initial appointment will be made on a fixed-term contract, with a housing compensation of 600,000 RMB upon fulfillment of the contract requirements. Re-engagement thereafter is subject to mutual agreement.

APPLICATION

Please submit completed application form, CV and cover letter via email to rsc@nbu.edu.cn. Application forms can be downloaded from http://www.nbu.edu.cn/shizi Recruitment stays open until positions are filled unless otherwise specified. Please visit http://rsc.nbu.edu.cn for more details.

INFORMATION ON POSITIONS

SCHOOL OF MARINE SCIENCE
Professor/Associate Professor
Marine Biotechnology/ Marine Sciences/Medicinal Chemistry/ Ocean Engineering/Marine planning and remote sensing /Sea port and environmental ecology/Marine Geographic Information Science

Faculty of Electrical Engineering and Computer Science
Professor/ Associate Professor Wireless Communications / New Generation Communication Networks / Underwater Acoustic Communication / Multimedia Information Processing / Embedded Systems / Integrated Circuit Design / Electronic Design Automation / Database System / Big Data Processing / Software and Theory / Information Security / Mobile Computing / Graphics and Image processing / Power system and Its Automation / Power Electronics / Pattern recognition and Intelligent system / Sensor and Intelligent detection

FACULTY OF SCIENCE
Professor/Associate Professor/Assistant Professor
Condensed Matter Physics/Microelectronics/Optoelectronics/Solar Cell
Department of Mathematics/Computational Mathematics/Probability and Statistics / Financial Mathematics

FACULTY OF MECHANICAL ENGINEERING AND MECHANICS
Professor/Associate Professor/Assistant Professor
Mechanics/Vehicle Engineering/ Mechanical Engineering/Industrial Design

INTERNATIONAL COLLEGE
Professor/Assistant Professor
Accounting

SCHOOL OF MATERIALS SCIENCE AND CHEMICAL ENGINEERING
Assistant/Associate/Full Professors
Polymer Science/Chemical Engineering/Materials Science

SCHOOL OF LAW
Professor/Associate Professor/Assistant Professor
Criminal Law/Criminal Procedure Law/Civil Procedure Law/Civil Law/Electronic Commerce Law

COLLEGE OF TEACHER EDUCATION
Professor/Associate Professor/Assistant Professor
Curriculum and Teaching Methodology/Higher Education/Preschool Education/Educational Economy and Management/Cognitive Psychology/Educational Psychology/Clinical Psychology/School Psychology/Experimental Psychology/Personality Psychology/Social Psychology/Management Psychology/Computer Graphics and Digital Image Processing/Electronic Music, Game Development/3D Animation and Game Development

FACULTY OF MARITIME AND TRANSPORTATION
Professor/Associate Professor/Assistant Professor
Department of Logistics and Transportation/Department ofMaritime Technology/Department of Marine Engineering/Department of Naval Architecture and Ocean Engineering

FACULTY OF PHYSICAL EDUCATION
Assistant Professor
Sport Management/Sport Sociology/Human Movement/Sport Training/Sport Physiology

MEDICAL SCHOOL
Professor / Associate Professor
Mechanism And Prevention Of Alzheimer's Disease; Oncology; Genetics; Human Anatomy; Histology And Embryology; Cell Biology; Immunology; Microbiology; Parasitology; Analytical Chemistry; Occupational And Environmental Health; Toxicology

COLLEGE OF ARTS
Professor / Associate Professor / Assistant Professor
Performance Of All Areas, Musicology (Ethnomusicology), Music Composition
Art Design Of All Areas, Digital Arts, Fine Art, Art History,Art/Music Industry, Art/Music Therapy

SCHOOL OF ARCHITECTURE, CIVIL ENGINEERING AND ENVIRONMENT
Professor /Associate Professor /Assistant Professor
Architecture Design/Urban Design/ Urban Planning/Architecture Technology/Human Geography/ Physical Geography/Cartography and Geographic Information System/Environmental Technology/Applied Environmental Microbiology/ Civil Engineering/Engineering Management

ScienceCareers online @sciencecareers.org
FOCUS ON CHINA



西安电子科技大学
XIDIAN UNIVERSITY

Xidian University sincerely invites outstanding talents in all related areas

Xidian University, located in the world famous historical and cultural city of Xi'an,is one of China's key universities under the direct supervision of Chinese Ministry of Education. While the university excels in information and electronic technology, it also develops well in the fields of engineering science, management and humanities. At present, the university has more than 30,000 undergraduates, including over 1,700 PhD candidates and 8,900 master candidates.

Xidian University is one of the first universities in mainland China establishing fields such as information theory, information system engineering, radar, microwave techniques & antenna, and mechatronics, opening up IT disciplines in mainland China, having trained more than 160,000 high-level talents in the electronic information field.The university has been committed to system development, the tackling of technological problems, engineering research and development, created many firsts in Chinese engineering information and other fields, made significant contributions to the national informatization and the modernization of national defense. Xidian University has 4 State Key Laboratories and 22 Provincial/Ministerial Key Laboratories and has established long-term strategic cooperative partnerships with more than 10 research institutes, research centers and enterprises as well as setting up joint laboratories with 46 multinational companies. For years, the university has undertaken more than 1600 key and major projects and scientific research tasks and produced a number of landmark research results. Entering the new century, the university has obtained 15 national science and technology awards. In 2013, the university, as the first organization, won the second prize of National Natural Science Award, the second prize of National Technological Invention Award and the second prize of National Technological Progress Award, becoming the only organization which wins three national awards in the information field. After "211 Project"、"985 Project", in the third national higher education strategic project—Promotion Plan Innovation Capabilities of Universities (2011 Project), facing the national strategic demand and task, Xidian university takes the lead in setting up Collaborative Innovation Center of Information Sensing Technology, and ranking first of industrial class in year 2014 in the national "2011 Project" comprehensive expert consultations.

How to Apply

Interested applicants should send resume with recent achievements of scientific research projects, publications, awards and related supporting materials via email, fax and other ways to relevant colleges or Human Resources Department of Xidian University.

Address : 266 Xinglong Section of Xifeng Road, Xi'an, China
Talent Development Division / Li,Ming Tel : 86 - 29 - 8189 1796
Talent Recruitment Division / Jiang, Lili Tel : 86 - 29 - 8189 1773
Fax : 86 - 29 - 8189 1796
Email : rckfk@xidian.edu.cn rcb@xidian.edu.cn

常州大学
CHANGZHOU UNIVERSITY

Changzhou University Overseas High-end Staff Recruitment for 2015



Changzhou University (CZU) is located in Changzhou, a famous historical and cultural city with highly developed economy. It is to the south of Yangtze River, Jiangsu Province, China. Founded in 1978, Changzhou University, authorized by the State Council, was officially named as Jiangsu Institute of Chemical Technology in 1981. In 1992, it became attached to China Petrochemical Corporation Group (originally China Petrochemical Corporation) and it was thus renamed as Jiangsu Institute of Petrochemical Technology, which was later changed to Jiangsu Industrial Institute. In 2010, the University was approved by the Ministry of Education to be named as Changzhou University. In 2011, the Government of Jiangsu Province signed an agreement with the China National Petroleum Corporation (CNPC) and the Sinopec Group Company to co-develop Changzhou University.

The University delivers progressive courses in applied sciences with industrial application, as well as courses in the humanities, art, languages, technology, and agriculture and food technologies, with well-integrated 'enterprises, universities and research institutions' across a beautiful and well-resourced campus.

In the past 36 years, the university has developed into an institution of higher learning strong in engineering, science and management enhanced by other disciplines. The university has one Materials and Engineering Science PhD program; 10 masters degree programs on the first level disciplines; 42 masters degree programs on the second level disciplines and seven engineering masters degree programs. The University has sixty-five undergraduate programs covering engineering, science, manage- ment, economics, literature, laws and arts, as well as medical, agricultural and education sciences.Changzhou University is offering job opportunities in areas below:

Chemical engineeringFood science
Material sciences Bio-pharmaceuticalsMachinery engineering
Environmental engineering
Rail transportation
Electronic and information engineering
Economic management
Law, languages and fine arts
Other unmentioned new and interdisciplinary areas

Discipline/academic leaders and Ph.D holders are welcome to join Changzhou University.

Applications can be sent to rsczp@cczu.edu.cn. Please put the subject of your email in the following order:

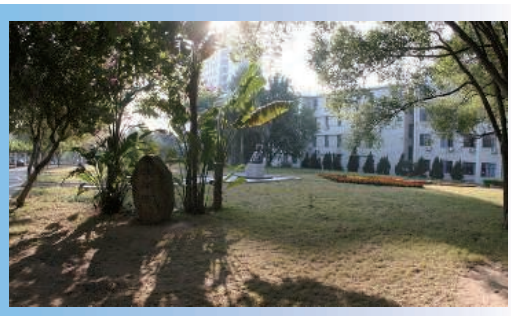
Name + area to be applied for + your final degree/ professional title.

For more information, you can visit our website http://rsc.cczu.edu.cn/.

For further details you can also contact Ms. Zhu, Mr Tian or Mr Zhou via 0086-519-86330024 or 0086-519-86330568.

HR office, 514 Wenzheng Building, Changzhou University, No.1 Gehu Road, Changzhou, Jiangsu, 213164, P.R. China.





華僑大學
HUAQIAO UNIVERSITY

Faculty Positions Available in Huaqiao University, Fujian, China

Founded in 1960, Huaqiao University is subordinate to the Overseas Chinese Affairs Office of the State Council of the People's Republic of China. It is now China's major base for overseas Chinese language and culture education. The University has campuses both in Xiamen and Quanzhou. Currently the university has 29 schools with 20 doctoral programs, 111 master degree programs and 84 bachelors programs. Besides, the university boasts 1 national key discipline, 40 key disciplines of provincial or ministerial level, 3 Engineering Research Centers of National Ministry of Education and 5 centers for postdoctoral mobile research stations. The student population totals up to more than 30000, of which more than 4000 overseas students come from 30 countries or regions including Hong Kong, Macau and Taiwan. Huaqiao University has become one of the universities that have the largest number of overseas students, and is committed to a high-level university with a solid foundation, distinctive features and high reputations from home and abroad. With this international recruitment, Huaqiao University invites applications for tenure track positions at the Professor, Associate and Assistant professor level all over the world.







Disciplines for Recruitment

Mechanical Engineering, Material Science and Engineering, Chemical Engineering and Technology, Electronic Science and Technology, Information and Communication Engineering, Computer Science and Engineering, Software Engineering, Civil Engineering, Geotechnical Engineering, Architecture, Urban and Rural Planning, Biomedical Engineering, Mathematics, Statistics, Applied Economics, Business Administration, Tourism Administration, Chinese Language and Culture Education, International Relations, Linguistics and Applied Linguistics in Foreign Languages.

Requirements & Qualifications

An applicant should possess an oversea PHD degree and show great interest in teaching and research. We are seeking applicants with exceptional records of academic achievements and give priority to those who currently hold professorship or associate professorship or other corresponding positions in well-known universities.

Benefits

Our University is proud of its highly supportive environment for researchers. The university will provide new faculty with high competitive salary, generous start-up funding and excellent research laboratory space, besides insurances, relocation fee, housing and other benefits.

Successful applicants will be granted the privilege of applying for Thousand Talents Program, Changjing Scholars Program, Minjiang Scholars Program and Thousand Talents Program for the Young. If successfully selected, the university will provide supporting treatment according to related regulations.

Contact us

Please submit materials including a cover letter, a CV and a brief description of future plan to rsc@hqu.edu.cn / hqursc@126.com with "Faculty Application from Science" in the title.

Contact Person: Li Meng, Wang Ke

Tel: 86-592-6161057, 86-595-22692370.

For more information, please visit the university's Human Resource Department website: http://rsc.hqu.edu.cn



Shanghai Maritime University seeks talents at home and from abroad

Shanghai Maritime University (SMU) is a multi-disciplinary university that encompasses such areas as engineering, management, economics, law, liberal arts, and science, with a special emphasis on shipping technology, economics and management. Chinese maritime education originated at Shanghai and grew out of the Shipping Section of Shanghai Industrial College founded in 1909 (towards the end of the Qing Dynasty). SMU was established by the Ministry of Communications in 1959. According to the university layout adjustment in Shanghai, and for better service of construction of Shanghai international shipping center, Lingang new campus was inaugurated in 2008.

Chinese maritime education originated at Shanghai and grew out of the Shipping Section of Shanghai Industrial College founded in 1909 (towards the end of the Qing Dynasty). SMU was established by the Ministry of Communications in 1959. According to the university layout adjustment in Shanghai, and for better service of construction of Shanghai international shipping center, Lingang new campus inaugurated in 2008.

At present the university runs 2 post-doctoral research stations, 19 doctoral programs, 59 master's degree programs, 45 bachelor's degree programs, and 12 associate degree programs. SMU boasts 9 provincial and municipal key research institutes and laboratories, 1 state-level key discipline, 5 state-level specialties with special features, 9 ministerial or municipal key disciplines, and 17 Shanghai municipal "educational heights". The SMU owns an aquatic training center, the Yufeng Ship, a 10-thousand-ton container ship for teaching and internship training, and a newly-built 48-thousand-ton Yangtze handymax bulk carrier for teaching and internship training ship, named Yuming ship.

In the MOE evaluation of undergraduate education in 2004, SMU was awarded an "A"(Excellent). Then in 2006 the SMU English-language specialty won another "A" in the MOE assessment. The university has won multiple awards for research including national awards for science and technology projects and awards for scientific and technological progress above the municipal and ministerial level in 2013.

SMU presently consists of Merchant Marine College, College of Transport and Communications, School of Economics & Management, College of Logistics Engineering (Sino-Dutch Mechanical and Electronic Engineering College), Law School, College of Information Engineering, College of ForeignLanguages, College of Ocean Science and Engineering, College of Arts and Sciences,Scientific Research Academy, etc. SMU has over 20,000 full-time students, including 17,000 undergraduates and over 3,000 graduate students. Of the over 1,000 full-time teachers, 151 are professors and 58 percent hold a doctorate degree. Over the past decades the university has produced specialists of various types at various levels for the country's shipping industry. The graduated students are employed in shipping companies, port enterprises and government institutions.Deservedly, SMU has been honored as a "cradle ofinternational shipping specialists".

The SMU established China (Shanghai) Institute of FTA Supply Chain and Shanghai Advanced Institute of International Shipping in 2013. The China (Shanghai) Institute of FTA Supply Chain make the combination of FTA construction and supply chain study, to boost the level of FTA industrial chain, accelerate the transformation from cargo trade to service trade, and promote the reform of government regulation function.

The Shanghai Advanced Institute of International Shipping adopts the advanced international operation mode and shares resource with other excellent education institutions, to build a domestic leading and renowned internationally shipping finance education base with high-end influential talents.

SMU is looking to further implement the strategy of strengthening the university with talent, and aims to recruit more high-level talents, ensure outstanding personnel qualifications and improve the development of academic programs and faculty. SMU has devised three plans for talent recruitment, including "Project Pilot", "Project Voyage" and "Project Sail".

I. Major Fields of Recruitment

Communication and Transportation Engineering, Management Science and Engineering, Naval Architecture and Marine Engineering, Transportation Strategy and Planning, Transportation



Management Modernization,Vehicle Operation Engineering, Logistics Management and Engineering, Logistics Equipment Safety Engineering, Port Power Transmission and Control Engineering, Power Electronics and Power Transmission, Electrical Engineering and Automation, New Energy, Mechanical Design Manufacture and Automation, Maritime Infor- mation and Control, Navigation Intelligent Decision-making and Simulation Technology, Navigation Environment Safety System, Vessel and Port Control Technology, System Simulation andControl Research, Marine Engineering, Vessel Refrigerated Transport Research and Moder- nization, Marine Engineering Management, Industrial Economics, International Law (Maritime Law), etc.

Institute of China (Shanghai) Pilot Free Trade Zone Supply Chain:
http://www.shmtu.edu.cn/sites/shmtu.edu.cn/files/cifsc-zp.htm

II. Qualifications

1.Doctorate degree received at renowned universities overseas;

2.Experience in prospective and innovative scientific research, publication of high-quality academic papers in major academic journals of relevant fields or prominent achievement in the field of engineering, and the potential of acting as the academic or technology leader in field of expertise;

3.Priority given to those with 2+ years of work experience at renowned universities or research institutions overseas.

III. Compensation

Based on the academic level evaluation of the applicant by the recruitment committee, for outstanding talents, SMU will offer a negotiable agreement of a competitive yearly salary, provide sufficient scientific research funds and a certain amount of housing subsidy. In addition, for particularly outstanding candidates, SMU will provide housing based on relevant provisions of the university.

IV. Materials for Application

1.Curriculum Vitae (including continuous education and work experience from undergraduate education to the time of application, publications, research projects, certification of awards, etc.);

2.Overview on academic success (roughly 500 words);

3. Major ideas for work if hired;

V. Contact Information

Address: Human Resources Department, Shanghai Maritime University, No. 1550 Haigang Avenue, Shanghai
Zip Code: 201306
Contact: Huafeng Wu, Jie Li
E-mail:hfwu@shmtu.edu.cn,
jieli@shmtu.edu.cn
Tel:021-38284233;021-38284232
Fax:021-38284238
Website:http://www.shmtu.edu.cn;
http://hr.shmtu.edu.cn/

上海海事大學
SHANGHAI MARITIME UNIVERSITY



New Recruiting Plan with Decent Salary: Young Talent Support Plan of Xi'an Jiaotong University

Xi'an Jiaotong University, a national key university under the direct jurisdiction of the Ministry of Education, is one of the country's oldest institutions of higher learning. Founded in Shanghai in 1896 as Nanyang Public, it was renamed Jiaotong University in 1921. In 1956, at the direction of the State Council, the university was moved to Xi'an, and renamed Xi'an Jiaotong University (XJTU). It was included in the first group of universities to benefit from Project 211 and Project 985, with the goal of becoming a world class university. Being a comprehensive research university and offering programs in ten areas—science, engineering, medicine, economics, management, art, law, philosophy, education, etc—with a major emphasis on science and engineering, Xi'an Jiaotong University is currently recruiting young scholars with outstanding performance to make further breakthroughs in its basic science frontline and interdisciplinary research. Since October 2014, The University has implemented "Young Talent Support Plan" with the aim of bringing together a number of emerging scholars who have demonstrated remarkable academic potential. Successful applicants will be offered with great support and a series of favorable policies in career development and personal well-being, including academic teaming, research platform, housing and children's education. We strive to create an environment that enables young scholars to quickly grow into leaders and pioneers in their academic fields. We cordially welcome all to join our family.

Program details are as follows:

1) Basic conditions, requirements and fields of recruiting

i.The program is open to all interested individuals, domestic and overseas. Successful applicants shall take full-time professorship in the University.

ii.The applicant is required to have high academic moral standards, be a lawful citizen and have a Ph.D degree. It is desired that the applicant be under the age of 40. For those with extraordinary excellence, the age limit is 45.

iii.The applicant should have made outstanding academic achievements in his/her field, being among the top talents in his/her contemporaries and have been well recognized by academic peers in the same field.

iv.The applicant should have a broad academic vision, great creativity, and potential for growing into leaders in his/her field.

v.Fields of recruiting:

Energy&Environment:
Power Engineering and Engineering Thermophysics, Nuclear Power and Nuclear Technology, Environmental Science and Engineering, Electrical Engineering, Chemical Engineering and Technology, Civil Engineering, Architecture, Municipal Planning.

Mechanics, Aerospace & Material:
Mechanical Engineering, Instrument Science and Technology, Mechanics, Aeronautical and Astronautical Science and Technology, Materials Science and Engineering.

Sciences:
Mathematics, Physics, Statistics, Chemistry.

Electronics&Information:
Electronics Science and Technology, Information and Communication Engineering, Control Science and Engineering, Computer science and Technology, Software Engineering.



Medicine&Life Sciences:
Basic Medicine, Clinical Medicine, Public Health and Preventive Medicine, Pharmaceutical Science, Stomatology, Chinese and Western Integrative Medicine, Nursing,Biology, Biomedical Engineering.

Humanities&Social Science:
Philosophy, Marxist Theory, Law, Sociology, Education, Science of Physical Culture and Sports, Foreign Languages and Literatures, Journalism and Communication, Design.

Management&Economics:
Management Science and Engineering, Business Administration, Public Management, Applied Economics, Theoretical Economics.

Interdisciplinary fields.

2) Application materials (all materials must be completed and submitted)

i.Contact specific schools or departments for detailed information and complete Application Form for the "Young Talent Support Plan"of Xian Jiaotong University.

ii.Submit related materials that demonstrate the applicant's academic achievements.

iii.Three to five recommendation letters from famous overseas experts in the same field. Selection is performed by Human Resources Department based on application materials and interviews.

3) Salary, benefits and other supports

i.Annual salary for each successful applicant will be RMB350,000 (before tax); annual salary will be RMB450,000 (before tax) for each successful applicant who enters Changjiang Scholars Program or who is awarded the National Science Fund for Distinguished Young Scholars. Annual salary for those with extraordinary excellence is negotiable.

ii.The School (the Discipline) will offer sufficient research resources and matching starting-up research funds for successful applicants.

iii.Successful applicants are qualified for supervising Master and Ph.D candidates.

iv.The University will offer high-quality resources for children's education.

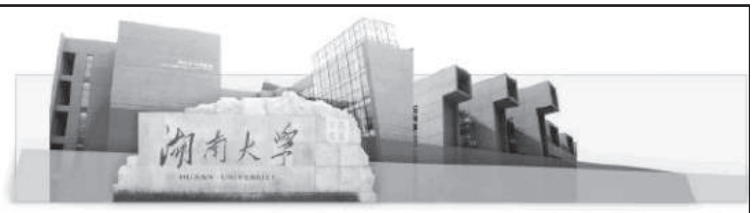
v.Successful applicants can rent or buy University housing; the University also offers a one-time settling-in allowance of RMB350,000.

The Program will be long effective. We warmly welcome your inquiry and application!

Contact: Mr. Dong Liang, Ms. Ying Liu
Address:
Human Resources Department,
Xi'an Jiaotong University #28 Xianning West Rd.
Xi'an, Shaanxi Province,China 710049
Tel: +86 29 82668267 +86 29 82668975
Fax: +86 29 82663876
Email: liangdong@mail.xjtu.edu.cn;
lybobo@mail.xjtu.edu.cn
Website: http://www.xjtu.edu.cn/



湖南大學
HUNAN UNIVERSITY



Faculty Positions Available at Hunan University





Hunan University is situated at the foot of the picturesque Yuelu Hill and on the west bank of the rippling Xiang River in Changsha, one of the top 10 livable cities in China. It is a key university affiliated with China's Ministry of Education and is included in the state's "Project 211" and "Project 985" for priority investment and construction.

Hunan University, dating back to Yuelu Academy founded in 976 A.D. during the Song Dynasty, is reputed as "A well-known Millenarian Academy and a Famous Centennial University". Carrying on the good educational tradition formed over the past1,000-odd years, we now have a graduate school and 23 colleges, offering programs in 11 fields, i.e., philosophy, economics, law, education, literature, history, science, engineering, management, medicine and art. It also has 25 first-tier doctorate programs, 41 first-tier master programs and 20 professional programs. 14 programs are national key disciplines (second-tier) and there are also 25 post-doctoral research centers. The university has a faculty of more than 1900, 1200 of whom are professors and associate professors. There are over 20,700 undergraduates and more than 14,000 graduate students.

Hunan University is a modern research university. We are striving to build Hunan University into one of China's first-class comprehensive, open and research-intensive universities with considerable influence in the world. We are recruiting highly-qualified and motivated faculty members and researchers in disciplines of Mathematics, Physics, Biology, Environmental Science, Material Science and Chemistry with negotiable salary and start-up research fund.

Please submit your detailed CV to Human Resources Department at szgl@hnu.edu.cn.

All positions are open to applications. Please feel free to contact us if there is anyquestion.

Contact Persons: Ms. Wang Qian; Ms. He Min
Human Resources Department, Hunan University
Yuelushan, Changsha, Hunan 410082, P. R. China
E-mail: szgl@hnu.edu.cn
Tel: 86-731-8882-2723
Website: www.hnu.edu.cn

西南大學

Faculty Positions Available in Southwest University, Chongqing, China

Southwest University invites applications from home and abroad for faculty positions at any level–assistant to full professors in ALL areas. Candidates must have a doctor's degree and exceptional research accomplishments in relevant research fields. Competitive packages including set-up and research funds as well as attractive salaries and benefits will be provided to all successful candidates. Distinguished candidates who have outstanding track record in research area and already got professor position abroad could be recruited through the national Recruitment Program of Young Experts or the national "One Thousand Talents Program". Successful applicants will be offered great support and a series of favorable policies in career development and personal well-being, including academic teaming, research platform, housing and children's education.

Southwest University is a comprehensive and national key university of the "211" project directly under the Ministry of Education. It is located in Chongqing. The university currently comprises 55 undergraduate programs, 46 master's programs, 21 doctor's programs, 22 postdoctoral research stations, 3 national key disciplines, 3 key disciplines with national support, 37 provincial key disciplines, 1 state key laboratory, and 10 key laboratories of Ministry of Education or Agriculture. Among its 2,700 faculty and staff members, there are 1 member of the Chinese Academy of Sciences, 2 members of Chinese Academy of Engineering,8 professors in National "One Thousand Talents Program", 7 "Chang Jiang Scholars", 2 recipients of National Science Fund for Distinguished Young Scholars, 2 chief scientists of National "973" research Projects, and 13 professors in "BaiQianWan Talents Program". The university got more than 700 national research projects and the research projects and fundsincreased by33% per year in the past five years.

Further information is available at http://renshi.swu.edu.cn/rcgzbgs/ or from the address: The Talents Recruitment Office, Southwest University, Beibei, Chongqing 400715, P. R. China. Please kindly send applications or nominations in the form of an application letter enclosing a current CV to rencai@swu.edu.cn.

Job Vacancies in China's Universities

赛尔互联 CERNET

Science Careers
From the journal *Science* AAAS

赛尔互联：《科学》在中国大陆高校人才引进服务独家合作伙伴
CER is Science's exclusive agent for recruitment advertisement service in mainland China universities and colleges

China's Rapid Development
—— More Opportunities

- **South China University of Technology** (Shenzhen, China)
  DeanPositions offered: The School of Mathematics, The School of Physics and Optoelectronics, and The School of Electronic and Information Engineering.
- **Jiangsu University** (Jiangsu, China)
  SearchingTalents for following disciplines: Mechanical Engineering/ Environmental Science and Engineering/Computer Science and Technology/Chemistry/ Basic Medical/ Fine Arts...
- **Beijing Institute of Technology** (Beijing, China)
  Open Senior Faculty Positions offered: Equipment Science & Technology Engineering/ Mechanical Engineering/ Information and Communication Engineering/ Mathematics...
- **Xi' an Jiaotong University** (Xi' an, China)
  Faculty positions offered in the Joint School of Sustainable Development (JSSD) of Sustainable Energy, Sustainable Materials and Sustainable Systems.

For more details, visit http://www.eol.cn/

广告联系：赵佳 zhaojia@cernet.com +86 10 62603373

eol 中国教育在线 www.eol.cn 是赛尔互联旗下品牌



Global Talent Recruitment of Nanchang University

南昌大学 NANCHANG UNIVERSITY

Nanchang University (NCU), a Project 211 key university co-supervised by Jiangxi People's Government and Ministry of Education of PRC, is comprehensive where Arts, Science, Engineering, and Medicine infiltrate while study, research, and enterprise integrate.

It currently runs 112 undergraduate programs covering 12 disciplines, 8 primary disciplines which confer Ph.D. degrees, 45 primary disciplines which confer master degrees, and 11 post-doctoral research stations. In addition, NCU boasts 1 national key laboratory, 1 national engineering technology research center, 2 national key disciplines, and 1 key humanity and social science research base of Ministry of Education

In order to build NCU into a high-level comprehensive university with visible local features, we are looking forward to the application of talents home and abroad, including outstanding talents, disciplinary leading talents, disciplinary direction leaders, and outstanding doctors and postdocs.

I. Requirements and Qualifications

1.1.Basic Requirements

A. Applicants are expected to be healthy both in body and mind, rigorous in academic research, and strong in teamwork.

B. Applicants are supposed to be eligible for talent recruitment of NCU in terms of research interest, academic level, and education levels.

1.2. Specific Requirements for Talents of Different Levels

A. Outstanding talents are supposed to be internationally well-known professors or distinguished scholars, academically advanced in research achievements recognized by domestic and foreign experts in the related fields and professionally capable of cohering multi-disciplinary co-operation and promoting related disciplines to reach an internationally leading level.

B. Disciplinary leading talents are required to be under the age of 55 with a doctoral degree, associate professorship or above, and great academic achievements in related disciplines, competent to obtain and hold national key research projects and international key co-operation programs and to lead related discipline to reach an internationally leading level.

C. Disciplinary direction leaders are required to be under the age of 45 with a doctoral degree and assistant professorship or above in oversea universities or research institutes, internationally advanced in related research areas and capable to obtain and hold national research projects or international collaboration projects.

D. Outstanding doctors and postdocs are required to be under the age of 35, graduated from oversea universities or distinguished universities in China with at least one-year oversea study or research experience, and published series of academic papers in SCI cited journals as the first author or corresponding author.

II. Benefits and Supports

NCU will offer you a high salary, research start-up fund, and housing fund not lower than other universities of its kind in China.

III. Application Procedures

Applicants are encouraged to apply by emailing us a self-recommendation letter or expert-recommendation letter, your CV, and photocopies of your research projects, patents, published academic papers, and other research achievements and honors.

You can also contact us directly or log into NCU website to know more details. Our recruitment will start since its publication.

IV. Contact Information

Person to contact: Luo Jian gcrc@ncu.edu.cn
Wu Qiang rscrsk@ncu.edu.cn

Tel: 0791-83969076 0791-83969074
Fax: 0791-83969078
Mailing address: Personnel Division of NCU, 999 Xuefu Road, Nanchang, Jiangxi Province, PRC
Zip code: 330031
NCU URL: http://www.ncu.edu.cn
Personnel Division URL: http://rsc.ncu.edu.cn







South China University of Technology Calls for Young and Extraordinary Talents

South China University of Technology (SCUT), formerly known as the South China Institute of Technology, was established in 1952. SCUT is on the University lists of "211 Project" and "985 Project", sponsored by Ministry of Education. SCUT knows well the importance of talent to a university striving to become one of the best of its class in the nation and highly respected by the academic world. That is why we are pleased to invite you, being a person of strong record in teaching and research, to work at SCUT and help us achieve our goal.

I. Recruitment Program of Global Youth Experts (National Program)

1.Qualifications for the position

Applicants are expected to meet the following requirements:

1) Research experience in natural sciences or engineering, under 40 years of age;
2) Doctorate from an accredited overseas university , with at least three years of oversea working experience in research or education;
   Or have received a Doctorate from a recognized Chinese university, with at least five years of oversea working experience in research or education;
   Or experts at top overseas universities, research institutes or companies;
3) Academic leader or potential leader in areas of expertise;

Exception:

Fresh PhD graduates with strong research records and scientists with distinguished academic achievements could be considered as competitive candidates without the restraint of age or working experience.

The application will be evaluated by the program director and finalized by the university.

Once the university confirms the qualification, applicants should work fulltime.

2.Salary and Benefits

Other than the benefits provided by the national program, SCUT provides:

1) Research funding (within 5 years): humanities/societies, ¥600,000; economics/ business management, ¥800,000; science/engineering research, ¥1,000,000; applied science/ engineering, ¥1,500,000.
2) Annual Salary: ¥300,000 (before tax)
3) Housing benefit:
   ¥300,000 home acquisition allowance (before tax), ¥200,000 housing allowance (before tax); Or on-campus housing and ¥50,000 relocation expenses (monthly rent should be paid in line with university requirements)

II. Extraordinary Talent Recruitment Project (University Program)

1.Qualifications for the position

1) Associate professor or professor at top overseas university;
   Or research/technical experts at top overseas institutes or companies;
   Age under 50;
2) Or at least three years of oversea working experience in research or education,with strong academic background, age under 40;
3) Work fulltime

2.Salary and Benefits

1) Research and development funding (within 5 years): humanities/societies and economics/ business management, ¥300,000 - ¥1,500,000; science/engineering research, ¥500,000- ¥3,000,000; applied science/ engineering, ¥700,000 -¥6,000,000; technical development, ¥5,000,000 and above.
2) Housing benefit: ¥300,000 - ¥600,000 home acquisition fee (before tax), ¥150,000 -¥200,000 housing allowance (before tax);

Or on-campus housing and ¥50,000 relocation expenses (monthly rent should be paid in line with university requirement)

In order to assist qualified talents to adapt into the academic environment of SCUT, we will provide support in application for temporary/permanent residency, visa, health insurance and child education etc.

Contact information

Web Site: http://www.scut.edu.cn
Address:
Room 308, The Personnel Office, South China University of Technology
No.381, Wushan RD., Tianhe District, Guangzhou, P.R.China, 510640
Tel: +86-20-87110072
E-mail: rscrcb@scut.edu.cn
Contact Person: Mr. Wang, Ms. Wang



Positions to Prepare You and Your Talents for Action at the Sino-Dutch Biomedical and Information Engineering School, Northeastern University, Shenyang,China

School Profile

The Sino-Dutch Biomedical and Information Engineering School (BMIE) was jointly established in2005by Northeastern University (NEU) and the Eindhoven University of Technology (TU/e) in the Netherlands with active support from the Neusoft Medical and Philips Healthcare. BMIE is qualified for the granting of Bachelor's, Master's and Doctoral degree. The school has 55 faculty and staff, and a total of more than 600 current students.

The program of Biomedical Engineering was launched at NEU in 1996. BMIE was authorized to confer Master degree in 1998 and first-level discipline of doctoral degree in 2006. The subject was approved as a provincial key discipline in 2008. The application for the MOE Key Laboratory of Medical Image Computing was successfully accepted in 2009. The program of Biomedical Engineering was approved as a featured specialty of MOE in the same year. The school was approved as a State Engineering Center of Medical Imaging and Intelligent Analysis by MOE in 2013.

Job vacancies

1. Senior management:
   (1) Deputy Dean-Teaching (2) Deputy Dean-Scientific Research
2.Professional course teachers:
   (1) Professor (2) Associate professor (3) lecturer (4) tutor
   Institutes: Biomedical Imaging, Biomedical Electronics, Biomedical Informatics
   Professional directions: Biomedical Engineering, Computer Science

Please hit ***www.bmie.neu.edu.cn*** for detailed information.

Applications

Please forward your CV, reference letters (at least 2 professionals of relevant industry) credentials, qualifications and any other proof materials reflecting your teaching and research background to liangjh@bmie.neu.edu.cn, or make a phone call at +86 24 8368 7851 for any queries.

Faculty Position Available
at China University of Petroleum in Qingdao

China University of Petroleum (UPC) is a national key university directly affiliated with the Ministry of Education and a member of the "211 Project" and "985 innovation platform for preponderant discipline" universities。Honored as "the cradle of scientific and technological talents for petroleum industry", UPC is an important base of training high-level talents for the petroleum and petrochemical industry and has already developed into a multi-disciplinary, well-rounded university focusing on petroleum and engineering.

UPC now has two campuses in Shandong Province, one in Qingdao and one in Dongying. Qingdao Campus, the main and new campus, is located in Qingdao which enjoys a high reputation for its charming scenery and favorable climate.

We warmly welcome scientific and technological elites all over the world to join us and build our university into a national renowned, high-level, research-oriented institute with its petroleum-related disciplines reaching the world's first-class level.

UPC provides eligible talents with a good academic environment and excellent working and living conditions.

◎1.Tenured professor of well-known overseas universities or experts and scholars who have published a thesis as the original author in "Science" or "Nature" can be recruited as a high level professor at UPC， and will be offered 2 million RMB settling-in allowance or housing subsidy and 3-8 million RMB scientific research and academic funds after the evaluation.

◎2.Experts or scholars who have a high academic reputation and are qualified leader in certain disciplines can be recruited as a high level or regular professor， and will be offered 1 million RMB settling-in allowance or housing subsidy and 1-3 million RMB Scientific research and academic funds after the evaluation.

◎3.Youth talents who qualify for China's "Youth 1000 Talents projects" can be offered 0.5 million RMB annual salary, 1.5 million RMB settling-in allowance or housing subsidy and 1-3 million RMB scientific research and academic funds.

◎4.Youth talents who receive the position as postdoctoral, assistant professor, associate professor or have a PhD from a famous overseas university or research institution can be offered higher treatment than similar talents in China.

Further information is available at http:// rsc.upc.edu.cn

Contact us:

Tel: 0086-532-86981808; 86981806 E-mail: teacher@upc.edu.cn

Recruitment Notice
High-level Position Offered in JiangsuNormalUniversity







JiangsuNormalUniversity is located at the conjunction of Jiangsu, Shandong, Henna, and Anhui provinces, the center of Huaihai economic zone——Xuzhou. It is a co-funded university by Jiangsu Provincial Government and State Ministry of Education.

Jiangsu Normal University has 3 trial faculties, 22 academic schools, Continuing Education College, International College, and independent Kewen Institute. There are currently 80 bachelor's program, 26 first level master's stations, which cover 11 subject categories, 25,000 full time undergraduates, and over 3000 postgraduates.

Jiangsu Normal University warmly welcomes domestic and oversea high-level talents joining us.

Position Offered:

Statistics, Optical engineering, chemistry, biology, education, design, track transit, architecture

Distinguished Professor: 2+ years oversea experience after obtaining doctor's degree, deputy professor or above in oversea famous universities, institutes, renowned enterprises or financial institutions; Hosted key scientific research project, with many influential papers published in key academic journals, or having key technology, or holding major patent, or achieving great results in technology transformation, or in the leading position in the researching area, very influential in certain academic area world widely; Age below 50; full time at position in the contracting period (5 years).

Outstanding doctoral/postdoctoral degree holders: graduates from first-class universities, with abundant academic results and development potential in scientific research; having published many academic papers in domestic and oversea influential academic journals for the recent years, Hosted or acted as major participants of key scientific researching projects; good team cooperation ability; age below 40; full time at position in the contracting period (5 years).

Salary and Conditions:

Competitive salary and conditions are offered, abundant scientific research fund and experimental offices, especially in favor of constructing researching team. Salary and conditions are upon negotiations.

Application Guide:

Make sure to send a curriculumvitae and a brief description ofthe research interests.
Professor/Senior Researcher Employment Certificate
Doctoral Terminal Degree Certificate
Research publications, academic monographs, projects and awards in recent five years
Three recommendation letters

Contact:

E-mail:rsk@jsnu.edu.cn
Tel:+86-516-83500203 +86-516-83500434 (for international teachers) Resume-receiving email:http://rczp.jsnu.edu.cn

---

Guangdong University of Technology

Guangdong University of Technology (GDUT), Guangzhou, China, Invites Applications and Nominations for "Deans of five different schools"& "Top Talents " chair professors &"One-Hundred Talents" professors and associate professors

Guangdong University of Technology, located in Guangzhou, a beautiful city in southern China, is a key multi-disciplinary university of Guangdong Province with a history of over 50 years. In order to realize the aim of leap-forward development, GDUT continuously implements the high-level talents search programs including "Deans of five different schools"& "Top Talents "chair professors &"One-Hundred Talents" professors and associate professors.

1.The "School Deans" and the university's "One-Hundred Talents" Professors

The university plans to invite scholars, experts and entrepreneurs to be our "One-Hundred Talents"Professors, and to those who are competent to build first-class schools are invited to work as deans of the following schools:

(1)School of Electromechanical Engineering
(2)School of Automation
(3)School of Chemical Engineering and Light Industry
(4) School of Materials and Energy
(5) School of Management

The eligible candidates must have a doctoral degree(s) or have professor titles. The applicants for The "School Deans" Program shall have good capacity in organization and coordination competence to lead the school for further development. We offer competitive compensation package and benefits for the qualified candidates. All the benefits can be negotiated during the interview.

2. Top Talents Recruitment Plan

The university plans to invite top scholars, experts and entrepreneurs who are honored as following titles:

(1) Academician: The annual salary and housing allowance can be negotiated during the interview.

(2) The National outstanding youth fund gainers, The Yangtze River Scholars; The National Thousand Talents Plan experts, Guangdong Province leading talents, Guangdong Province Innovation team leader: The annual salary can be approximately from600,000RMB(USD100,000) to￥1000,000 RMB(USD160,000), housing allowance can be approximately from￥700,000 RMB(USD110,000) to ￥1000,000 RMB (USD160,000).

(3) The National Excellent youth fund gainers, the annual salary can be approximately ￥450,000 RMB(USD75,000) . The National Thousand Young Talents Plan experts, the annual salary can be approximately ￥ 400,000 RMB(USD65,000) , housing allowance can be up to￥700,000 RMB(USD115,000) The startup R&D funding for Top Talents can be up to ￥10,000,000 RMB (USD1600,000).

3. The University's "One-Hundred Young Talents" Associate Professors

This program is specially designed for young talents whose age is below 35, hold a doctoral degree(s), achieved satisfactory initial results. The university Provides competitive salary packages, the annual salary for the associate professors is over ￥200,000 RMB(USD35,000); Provide a temporary apartment and housing allowance; Startup R&D funding and laboratory space will also be offered.

4. Contact Information

All the documents and inquiries (including CV and work plans etc.) should be addressed to:

Contact person: Ms Ruiyin Yang Telephone: 86-20-39322509
Email: gdutyzzp@gdut.edu.cn
More information can be found on the university website:
www.gdut.edu.cn

江西师范大学
Jiangxi Normal University since 1940

Jiangxi Normal University Recruits Talents from globally

Jiangxi Normal University is located in Nanchang, the capital of Jiangxi province. It has been one of the members of state-province jointly construction and mid-western infrastructure projects since 2012.The university is rich in historical and cultural background and well-known for its reputation for natural abundance and heroic endeavor. At present, there are 3 campuses, at Lake Yaohu Campus, Lake Qingshanhu Campus and Gongqing City Campus. The University's fields of study include History, Philosophy, Economics, Management, Law, Physics, Engineering, Education and Fine Arts. Teacher training and non-teacher training co-exist on the three campuses. As its name implies, it is concerned with teacher training, but not to the exclusion of other fields of study. The University has a considerable influence on the political, economic, cultural and social development of Jiangxi province, and the provincial government has designated it a priority normal university at the provincial level.

For accelerating the course of faculties internationalization and promote the level of teaching staff, we now recruiting the high-ends talents from home and abroad.

I.Recruitment Disciplines

Foreign Languages and Literatures, Psychology, Mathematics, Physics, Geography, Chemistry, Chemistry Engineering and Technology, Computer Science and Technology, Biology, Theoretical Economics, Practical Economics, Management Science and Engineering, Business Administration and other newly-developing related interdisciplines.

II.Requirements and Qualifications

1.Leading personnel

Candidates should be the finalist of Chang Jiang Scholars Program and Recruitment Program of Global Experts of Central Organization Department. The scholars hold the post of professor or other equivalent position in renowned overseas universities or research institutions will be considered, as well as the senior positions or technique personnel at well-known enterprises. The age of candidates should under 55, and enjoys good health.

2.Academic Leader

Candidates should be the finalist of Young Overseas High-level Talents Introduction Plan, and hold the post of associate professor other equivalent position in renowned overseas universities. Other talents employed as JNU's academic leader of urgent need disciplines with doctorate acquired from renowned overseas universities will be considered as well. The age of candidates should under 45, and enjoys good health.

3.Outstanding young returnee doctors

Candidates should acquire the doctorate from renowned overseas universities, or the same level talents under the joint-supervision of domestic key university or research institution and renowned overseas universities who acquired the doctorate from overseas. The age of candidates should under 35, and enjoys good health.

III.Contact

All applicants please sent the curriculum vitae with working experience and performance, and 2-3 papers to jxsdrsc@jxnu.edu.cn

Address: No.99, Ziyang Rd, Nanchang, Jiangxi, Jiangxi Normal University, Yaohu campus

Tel: +86-791-88155787

Contact person:

Jiutian He(+86 13970904629)

Post code:

330022

Faculty Positions Available at Harbin Engineering University

Harbin Engineering University belongs to the first group of national "211 Project" universities, and also it enters the setting up of national "985 Project" frontier innovation platform projects. It is also a key university which contains graduate school, and China's "three sea one nuclear" (shipbuilding industry, naval equipment, marine engineering, nuclear energy applications) important personnel training fields and scientific research base.

We are hiring talented people whose majors are the following subjects: Naval Architecture and Ocean Engineering, Mechanics, Power Engineering and Engineering Thermophysics, Control Science and Engineering, Instrument Science and Technology, Computer Science and Technology, Mechanical Engineering, Information and Communication Engineering, Management Science and Engineering, Material Science and Engineering, Nuclear Science and Technology, Optical Engineering.

1.Outstanding leading talents: talents should have the leading abilities of surpassing or maintaining the international advanced level of the related disciplines, including professors of the famous overseas universities, chief scientists of national "973 project", leading talents of national "people plan", long-term experts and candidates of national " recruitment plan", chair professors of "Cheung Kong Scholars Program", winners of "National Science Fund for Distinguished Young Scholars" and national teaching masters etc.

2. Young leading talents: talents should have academic influences in the forefront of this field at home and abroad. They should aim at being the candidates of "Cheung Kong Scholars Program", winning the "National Science Fund for Distinguished Young Scholars" and being awarded as the national teaching masters, etc.

3. Young top talents: talents should be the top and leading young figure of his working and innovation. They should aim at being the candidates of "Recruitment plan of Global Young Experts", "Supported by Program for the Top Young Talents", winning the "National Science Fund f or Distinguished Young Scholars" and "Program for New Century Excellent Talents in University" awarded by the Ministry of education.

4. Excellent young teachers: talents should have the PHD in overseas high-level universities; or talents with PHD who have the experience of more than two years' teaching and research in overseas high-level universities or scientific research institutions; talents should have the potential of good academic foundation and young top talents.

Our university will provide the talents different competitive annual salary, innovative technology and research platform, academic innovative stuff, safe and peaceful environment.

Contacts: Tan Sichao Tel: 86-0451-82519965
Wu Lei Tel: 86-0451-82589527

E-mail: rencai@hrbeu.edu.cn

website: http://www.hrbeu.edu.cn/; http://hr.hrbeu.edu.cn/

華東交通大學
EAST CHINA JIAOTONG UNIVERSITY

Golobal Talent Recruitment of East China Jiaotong University

East China Jiaotong University (ECJTU) is an education and research university focusing on engineering and technology. It is located in the heroic city of Nanchang, which has enjoyed a reputation for being richly endowed with natural resources and talents since ancient times. In 1971, the State Council, Central Military Commission approved to combine the School of Locomotive of Shanghai Jiaotong University with the School of Railway Engineering of Tongji University to become Shanghai Railway Institute and it was moved to Jiangxi Province, renamed as East China Jiaotong University and had been affiliated to the Railway Ministry of China . In March of 2000, ECJTU was reformed to be co-sponsored by both central and provincial governments(mainly by the local government). It is listed as a key university entitled to the funding support from the government. It is also an organization being constructed for PHD degree program.

Specific Requirements on All Levels of Talents

1.Leading talents: Internationally well-known professors or distinguished scholars recognized in the same field, having high academic attainments, taking the leading position internationally or in the same field in terms of academic level, obtaining internationally recognized important achievements, being capable of coagulating interdisciplinary cooperation and leading relevant disciplines up to internationally advanced level.

2.Academic leaders: possessing the doctorate and the title of associate professor or above in an overseas famous university or research institution, being generally 55 years old or below, having high academic attainments in a discipline, being among the best young and middle-aged scholars in this discipline field in terms of academic attainments, having obtained internationally advanced or domestic leading academic achievements, being capable of striving for and taking over national key research projects and important international cooperation projects, and being able to lead this discipline up to internationally advanced level

3.Academic backbones: possessing the overseas doctorate, being generally 45 years old or below, having the title of associate professor or above in an overseas university or research institution, being capable of mastering the development frontier of this discipline, having obtained a certain academic achievements, being able to apply for and take over international cooperation or national research projects as well as reach internationally advanced level in a certain research direction

4.Excellent doctoral graduates or post-doctoral outbound personnel: possessing the overseas doctorate or the doctorate from a domestic first-class university, having one year (or above) of experience in overseas study or cooperative research, being generally 35 years old or below, having issued several theses on the SCI Journal by the name of the first author or the corresponding author, being capable of applying for and taking over national scientific research projects.

Contact information:

Zhang yiyao 0791-87046604 13707093709

E-mail: rscecjtu@126.com

Address: NO.808 ShuangGang East Street

Changbei Economic and Technological Development Zone,Nanchang Jiangxi Province ,CHINA

Website: http://rsc.ecjtu.jx.cn/

华北电力大学
NORTH CHINA ELECTRIC POWER UNIVERSITY

Faculty Positions Available in North China Electric Power University, China

Established in 1958, located in Beijing and Baoding (main campus in Beijing), North China Electric Power University (NCEPU) is one of the top research-oriented universities in China, focusing on fundamental cutting edge research and high-level education, covering diverse areas of science, engineering, technology, humanities, economics, management and law. NCEPU, as a key university jointly constructed by the Ministry of Education and the University Council, an organization composed of seven super-large Chinese power enterprises, China Electricity Council and NCEPU, is affiliated with the Ministry of Education, officially listed as one of the "211 Project" and "985 Project" universities as well as a "Predominant Discipline Innovation Platform". It owns 1 national key laboratory, 2 state science and technology innovation platforms and 11 provincial and ministerial key laboratories. At present, the campus covers one million square meters in total, with about 3000 faculty and staff.

North China Electric Power University is on a clear path to grow as a world-class university in engineering and science fields. As part of university's further pursuit for excellence in research and education, we have expanded a global search for the best research talents to join us.

North China Electric Power University invites applications for full-time Professors, Associate Professors and excellent scientists. Preference will be given to candidates whose research emphasis demonstrates the potential to complement and advance our existing research strengths on electric power, clean energy, energy power, mechanical engineering, automation, computer science, chemistry, economics, management, mathematics and physics. Successful candidates will be offered competitive salaries, appropriate positions and start-up funds.

Position Openings

- Position offered by the Recruitment Program of Global Experts (1000 Plan Professorship)
- Position offered by Chang Jiang Scholars Program
- Position offered by the Recruitment Program of Global Young Experts (1000 Plan Professorship for Young Talents)
- North China Electric Power University's Professors
- North China Electric Power University's Associate Professors
- North China Electric Power University's Teachers

Interested individuals should email his/her CV to ncepuscience@163.com, with "Faculty Application from Science" in the title. For more details, please visit our Talent Management Office's website at http://rcb.ncepu.edu.cn, or contact us by email ncepuscience@163.com, or telephone 86-010-61772414.

浙江工業大學
ZHEJIANG UNIVERSITY OF TECHNOLOGY

Faculty Positions
Zhejiang University of Technology
Calls for Talents from Home and Abroad

Zhejiang University of Technology（ZJUT）was established in 1953. Currently, it consists of 3 campuses, covering a total area of over 213 hectares, with a floor space of 840,000 square meters. It has 21 colleges and 2 departments with 3,358 faculty and staff, over 36,000 full-time students. Other important facts and figures are as follows: 2 academicians of CAE (Chinese Academy of Engineering), 3 shared academicians of CAE and CAS (Chinese Academy of Sciences), 4 scholars under the national Recruitment Program of Global Experts, 1 Chair Professors under Chang Jiang Scholars Program, 4 scholars awarded with National Science Fund for Distinguished Young Scholars. The university is one of the first fourteen leading universities in the '2011' Project (the 3rd national higher education project, focused on collaborative innovation).

ZJUT now seeks for outstanding scholars and technicians in the following and relevant fields:

* Chemical Engineering & Techniques, Material Science & Engineering, Environmental Science & Engineering, Pharmaceutical Engineering, Biological Engineering, Civil Engineering, Food Science & Engineering;
* Mechanical Engineering, Control Science & Engineering, Electronic Science & Technology, Computer Science & Technology, Information & Communication Engineering, Software Engineering, Power Engineering, Shipping & Marine Engineering;
* Mathematics, Physics, Optic Engineering, Chemistry, Marine Sciences;
* Applied Economics, Theoretical Economics, Business Administration (Management), Public Administration, Management Science & Engineering;
* Architecture, Design, Urban & Rural Planning;
* Education, Journalism & Communication, Language & Literature, Law.



Positions Available

* Position offered by Recruitment Program of Global Experts (Also known as the "Thousand Talent Program")
* "Qianjiang" Special-Term Scholar Position offered by Zhejiang Province
* "Yunhe" Special-Term Scholar
* Professor
* Lecturer

Salary, Benefit and Support

Zhejiang University of Technology will provide the state-of-the-art facilities, exciting research environment, start-up funding, housing subsidy, competitive salary and other employee benefits. The exact amount differs from person to person according to agreement.

Contact Us

For more information visit www.zjut.edu.cn.

To apply, please send your CV and relevant materials to rcyj@zjut.edu .cn or contact us through +86-571-88320433.

上海交通大學
SHANGHAI JIAO TONG UNIVERSITY

Multiple Faculty Positions in Ultrafast Sciences / Electron Microscopy

Shanghai Jiao Tong University, Shanghai, China

Shanghai Jiao Tong University is establishing a world-class research Center for Ultrafast Sciences in physics, chemistry, materials and biology with integration of both experimental and theoretical efforts. The Center will provide state-of-the-art ultrafast laser spectroscopy, electron diffraction and microscopy, and excellent multi-disciplinary environment for cutting-edge research.

The Center invites applications for a number of open-rank faculty positions to begin in the fall of 2015 (or earlier). Applicants should have a Ph.D. degree, preferably with postdoctoral experience in a closely related field, and a strong record of research accomplishments. Previous experience in electron microscopy and ultrafast sciences is favorable. The successful candidates will be expected to develop world-class research programs and teach classes at both undergraduate and graduate levels. The University will provide attractive annual salary, start-up fund, housing allowance and other benefits.

All applicants should send a cover letter, a curriculum vitae with a publication list, a research proposal (3-4 pages) and a statement of teaching interest in a single pdf file by Jan. 31, 2015 to Dr. Jie Chen, Shanghai Jiao Tong University, Shanghai 200240, China through e-mail to: ultrafast@sjtu.edu.cn. Please also arrange three reference letters directly to the above e-mail address. Applications after the deadline could be reviewed until the positions are filled.

http://join.sjtu.edu.cn/

暨南大學
JINAN UNIVERSITY

Recruitment of Researchers for Cancer Institute of Jinan University

Jinan University is the first overseas Chinese university founded by the government in China, and the key comprehensive "211 Project" university directly under the Overseas Chinese Affairs Office of the State Council. It has six campuses which are distributed in Guangzhou, Shenzhen and Zhuhai, 15 post-doctoral research stations, four national key disciplines, eight key disciplines of the Overseas Chinese Affairs Office of the State Council, 20 first-level key disciplines of Guangdong Province and four second-level key disciplines of Guangdong Province.

Cancer Institute of Jinan University is the research institute directly under Jinan University, with fellow of Academy of Sciences at Abroad and Academy of Medical Sciences as honorary director and leader, has a research team consisting of Changjiang Scholars by the Ministry of Education, 973 Project's chief scientists, winner of National Outstanding Youth Fund and New Century Excellent Talents by Ministry of Education, and carries out innovative research on tumor molecular pharmacology, epigenetics, proteomic, molecular diagnostics, drug development, etc.

It now needs to recruits independent PI and postdoctoral researchers from China and other countries.

1.Requirements

Those who are from majors related to medical biology, have earned doctor's degree as well as have overseas study and work experience.

2.Recruitment Procedures

Those who meet the job requirements can make applications by email, and those who pass the preliminary examination will have an interview. The best candidates will be admitted.

3.Application Material

(1) Personal resume

(2) Copies of ID card (passport), diploma and vocational certificates

(3) Introduction to academic research and research programs

(4) Materials that can fully reflect academic levels, ability and achievements

4. Remuneration

Refer to the relevant documents of the school on talent introduction, and the specific is negotiable.

5. Contact E-mail

E-mail:jndxzls@163.com

天津科技大学

Faculty Positions Open to Talents from Home and Abroad

Tianjin University of Science and Technology（TUST） is a key comprehensive university in China and enjoys a high reputationin four preponderant disciplines, including food science and engineering, fermentation engineering, paper making and marine science and engineering. TUST, founded in 1958, is one of the first batch of Light Industry Institutes and a full-time state-run university which is jointly financed by the central government of China and the local government of Tianjin. The university is recruiting talents from home and abroad. Welcome to join us for your personal development and for the growth of the university.

1. Major Open Faculty Position

Food Science and Engineering	Biological Engineering/Biotechnology	Fermentation Engineering	Pulping and Papermaking
Mechanical Engineering	Chemical Engineering	Ocean Science and Engineering	Automation
Electronic Information	Computer Sciences	Packaging and Printing Engineering	Material Sciences
Law and Politics	Economy and Management	English and Japanese	Artistic Design

2. Talent Program

*TUST is especially interested in applicants who have working experience at major overseas universities and scientific research institutions and have been appointed as assistant professors and above.

*Applicants are supposed to receive a Ph.D from major universities either at home and abroad.

*The candidates have the access to Recruitment of Global Experts (Its subprograms include Short-term Program, Foreign Experts Program and Young Overseas High-level Talents Introduction Plan) supported by the Chinese government, Chair Professor of Cheung Kong Scholarship Program supported by Ministry of Education (MOE) as well as National Science Fund for Distinguished Young Scholars (NSFDYS).

*Besides that, outstanding faculty members have the opportunity to be admitted as the Chief Scientist of 973 Program initiated by the People's Republic of China, as the head of Project supported by the Funds for Creative Research Groups of China, and as the recipients of National High Level Personnel of Special Support Program (Its subprograms include Leading Talents and Young Top-notch Personnel).

*Furthermore, the candidates also have the access to Tianjin's "1000 Talent Plan" (including Short-term Program, Foreign Experts Program and Young Overseas High-level Talents Introduction Plan), The National Science Fund for Distinguished Young Scholars, Tianjin Distinguished Chair Professor, and Distinguished Scholars of Haihe Scholars as well as cultivating plan of Haihe Scholars supported by the university, etc.

3.Pay and benefits

For specifics of pay and benefits, please visit http://hr.tust.edu.cn/html/fagui/396.Html.

4. Contact us:

The applicants should submit a full CV, including a degree certificate, thesis, certificates of research, grants and awards via email to: person88@126.com or personnel@tust.edu.cn.
Contacts: Gou Junping, Wang Hongjun
Tel: 86-22-60273330 86-22-60271990
Add: Human Resources Office, No 1038 Dugu Nanlu, Hexi District, Tianjin, P.R.China
Postcode: 300222

中国民航大学

Faculty Positions Available at Civil Aviation University of China

http://www.cauc.edu.cn

Founded in September 1951, CAUC is located adjacent to Tianjin Binhai International Airport. It is an important base for cultivating professional talents and offering scientific and technical support for China's civil aviation industry.
CAUC is striving forth toward the goal of becoming "a teaching-and-research-oriented civil aviation university with nationwide prestige and international influence". We sincerely invite talents with lofty ideals to join CAUC!
More information available at http://www.cauc.edu.cn

I. Research Fields

Aeronautical & Astronautical Science and Technology, Traffic and Transportation Engineering, Control Science and Engineering, Safety Science and Engineering, Computer Science and Technology, Industrial and Commercial Management, Information & Communication Engineering, Aviation Law

II. Positions and Requirements

Position offered by Recruitment Program of Global Experts (1000 Plan Professorship)
Position offered by the Chang Jiang Scholars Program
Position offered to professors (associate professors) who currently work in high-level universities/institutes
Position offered by CAUC's Lantian Scholar Program of professors.
Note: Candidates for the above positions are supposed to be no more than 50 years old. Age limit could be extended in the urgently-needed disciplinarians.
Position offered by Recruitment Program of Global Young Experts (1000 Plan Professorship for Young Talents)
Position offered by CAUC's Lantian Scholar Program of Young Scholars
Note: Candidates for the above positions are supposed to be no more than 40 years old.

III. Salaries and Benefits

We will provide competitive salaries and welfares which include settling-in allowance (0.5-1 million RMB), rental allowance (1-1.5 million RMB), start-up fund for scientific research (1-10 million RMB), support in establishing scientific research platforms and groups. We can offer further or specific treatments that can be discussed personally.

IV. Application Materials:

Please send us by email a cover letter, complete CV, scanned diploma, certificate of degree and certification of R&D achievements, including statement of research projects, honor certificate of R&D achievements, patent certificate, a list and samples of publications.

V. Contact:

For further information, please contact our Human Resources Department:
Email: j_guo@cauc.edu.cn (Mr. Guo); lfyu@cauc.edu.cn (Mr. Yu); yai@cauc.edu.cn (Ms. Ai)
Website: www.cauc.edu.cn
Tel: 86-22-24092163 Fax: 86-22-24092163
Add: Human Resources Department of Civil Aviation University of China,
No. 2898 Jinbei Road, Dongli District, Tianjin, P.R. China
Post code: 300300

赛尔互联 CERNET
赛尔互联——科学杂志中国独家合作伙伴（限大陆高校招聘广告）
请访问 Sciencecareers.org/CER
获取全国各地联系信息
赛尔互联为您量身打造
海外人才引进解决方案
广告联系：赵佳 zhaojia@cernet.com
+86 10 62603373
中国教育在线 是赛尔互联旗下品牌

澳門大學
UNIVERSIDADE DE MACAU
UNIVERSITY OF MACAU

International Search for Chair Professors and Distinguished Professors

The University of Macau is a leading higher education institution in Macao, with English as its working language. In recent years, the University has made great progress in various areas, with increasing international recognition of its excellence in teaching, research and community service. The Times Higher Education World University Rankings 2014/2015 ranks the University among the world's top 300 universities. With the beautiful new campus (20 times larger than the old one) becoming fully operational, the launch of Asia's largest residential college system, the establishment of new faculties, and the increasing numbers of students and faculty members recruited from around the world, UM possesses great potential and provides exciting new possibilities for growth and development.

We plan to develop a strong team of top-notch scholars to help us realize our vision. Applications are therefore invited from those with excellent academic achievements in the following disciplines:

* Business and Management
* Education
* Law
* Liberal Arts and Humanities
* Social Sciences
* Mathematics, Sciences and Engineering
* Health Sciences
* Chinese Medicines
* Design

Remuneration and appointment rank offered will be competitive and commensurate with the successful applicants' academic qualification, current position and professional experience. The current local maximum income tax rate is 12% but is effectively around 5% - 7% after various discretionary exemptions.

For details about the above open positions and related information, please visit the following websites:

Job vacancy website: **http://www.umac.mo/vacancy**
University website: **http://www.umac.mo**
Macao government website: **http://www.gov.mo**

Applicants please state the specific disciplines applied for and quote the **reference no.: ACAD/PROF/09/2015** in their application and send it by email to the below address preferably on or before 12 December 2014. Review of applications will commence immediately and continue until the positions are filled. Applicants may consider their applications not successful if they were not invited for an interview within 3 months of application.

Human Resources Office
University of Macau
Av. da Universidade, Taipa, Macau, China
Website: https://isw.umac.mo/recruitment;
Email: senior.academic@umac.mo
Tel: +853 8822 8570; Fax: +853 8822 2412

The effective position and salary index are subject to the Personnel Statute of the University of Macau in force. The University of Macau reserves the right not to appoint a candidate. Applicants with less qualification and experience can be offered lower positions under special circumstances.

****Personal data provided by applicants will be kept confidential and used for recruitment purpose only****

** Under the equal condition of qualifications and experience, priority will be given to Macao permanent residents**

University of Macau - An ideal place to pursue your career
http://www.umac.mo

THE CHINESE UNIVERSITY OF HONG KONG

Applications are invited for:-

Department of Microbiology

The Department has a wide range of research facilities and access to a large comprehensive teaching hospital. The establishment provides a good environment for basic as well as clinical research and facilitates collaboration with other disciplines. Further information about the Department is available at http://www.cuhk.edu.hk/med/mic/.

(1) Professor / Associate Professor / Assistant Professor (Non-clinical)

(Ref. 1415/124(665)/2) (Closing date: January 12, 2015)

The Department invites applications for a Professorship / Associate Professorship / Assistant Professorship. Applicants should have (i) a PhD degree or equivalent; (ii) a strong research track record in the field of microbiology or virology; (iii) commitment to undergraduate teaching and postgraduate students supervision. Applicants for *Professorship / Associate Professorship* should also have a strong record on external funding support, international recognition and leadership. Experience in tumour virology, antibiotic resistance and microbial genomics will be advantageous, though not a pre-requisite. The appointee will (a) deliver teaching and related educational activities for undergraduate and postgraduate students; (b) supervise MPhil and PhD students; (c) apply for competitive research grants and related funding; and (d) conduct high-standard research projects independently and in collaboration with other parties. Appointment will normally be made on contract basis for up to three years initially commencing as soon as possible, which, subject to mutual agreement, may lead to longer-term appointment or substantiation later. Competitive start-up support will be provided.

(2) Research Assistant Professor

(Ref. 1415/125(665)/2) (Closing date: January 12, 2015)

The Department invites applications for a Research Assistant Professorship. Applicants should have (i) a PhD degree or equivalent; and (ii) a strong research track record in the field of microbiology or virology. Experience in tumour virology, microbial genomics and antibiotic resistance will be advantageous, though not a pre-requisite. The appointee will (a) apply for competitive research grants and related funding; and (b) conduct high-standard research projects independently and in collaboration with other parties. Appointment will initially be made on contract basis for up to three years commencing as soon as possible, renewable subject to mutual agreement.

Salary and Fringe Benefits

Salary will be highly competitive, commensurate with qualifications and experience. The University offers a comprehensive fringe benefit package, including medical care, plus a contract-end gratuity for appointments of two years or longer, and housing benefits for eligible appointees. Further information about the University and the general terms of service for appointments is available at http://www.per.cuhk.edu.hk. The terms mentioned herein are for reference only and are subject to revision by the University.

Application Procedure

Application forms are obtainable (a) at http://www.cuhk.edu.hk/personnel, or (b) in person/by mail with a stamped, self-addressed envelope from the Personnel Office, The Chinese University of Hong Kong, Shatin, Hong Kong, or (c) by fax polling at (852) 3943 1461. Please send the completed application form and/or full curriculum vitae, together with copies of qualification documents, a publication list and/or abstracts of selected published papers, and names, addresses and fax numbers/e-mail addresses of three referees to whom the applicants' consent has been given for their providing references (unless otherwise specified), to the Personnel Office by post or by fax to (852) 3943 1462 by the closing date. Please quote the reference number and mark 'Application – Confidential' on cover. The Personal Information Collection Statement will be provided upon request.

ScienceCareers online @sciencecareers.org

UNIVERSITY OF CAMBRIDGE

www.jobs.cam.ac.uk

Herchel Smith Postdoctoral Research Fellowships

Schools of the Biological and the Physical Sciences
£28,695 to £37,394

Funding is available for a maximum period of three years.

The Managers of the Herchel Smith Postdoctoral Research Fellowships Fund invite applications for these Fellowships within the fields of the Biological Sciences (particularly in fields at the interface with the physical sciences); Astrophysics; High-Energy Physics; and Pure Mathematics.

The Managers expect to offer four Fellowships in the Biological Sciences, three in the Physical Sciences, and one in Pure Mathematics.

The Fellowships are intended to enable a small number of outstanding and highly motivated scholars who have demonstrated a capacity for individual research to devote themselves exclusively to research at an early stage in their careers, to develop their academic potential, and to produce work of an international standard within their field.

All Fellowships are to be held from 1 October 2015 or otherwise by negotiation. Fellowships are available for between two and three years. Fellows will usually work in close collaboration with an established research group. Fellows are required to pursue their research on a full-time basis in Cambridge.

In accordance with Dr Smith's will, candidature is limited to candidates who have obtained their PhD degree, or equivalent, within the last three years at any university but normally excluding Cambridge and Harvard. Fellowships can only be offered to those with less than three years postdoctoral experience by the time the Fellowship is taken up. In addition to holding a PhD or equivalent, candidates should provide evidence of: outstanding academic excellence; a high quality and original research proposal; and evidence of their capabilities as an independent researcher.

The stipend will be on the University's Postdoctoral Research Associate scale, currently between £28,695 to £37,394 per annum, with an annual research allowance of £15,000 (£13,000 in Pure Mathematics).

Further details are available at http://www.herchelsmith.cam.ac.uk/fellowships/ or from the Secretary to the Fund Managers, Tel. +44 (0)1223 332283, e-mail duncan.mccallum@admin.cam.ac.uk

Please quote the following reference on your application and in any correspondence about this vacancy:

AK04854 for Fellowships in the Physical Sciences

AK04855 for Fellowships in the Biological Sciences

LF04858 for the Fellowship in Pure Mathematics

Closing date: 05 January 2015

The University values diversity and is committed to equality of opportunity.

The University has a responsibility to ensure that all employees are eligible to live and work in the UK.

University of Nevada, Reno

Department of Biology
Three tenure-track assistant professors

The Department of Biology at the University of Nevada Reno invites applications for tenure-track Assistant Professorships in **Organismal Animal Physiology, Metabolomics,** and **Disease Ecology** to begin July 1, 2015. The successful candidates will be expected to establish extramurally-funded research programs and to participate in undergraduate and graduate education. Startup packages for all three positions are highly competitive.

First, we seek an **organismal physiologist** preferably working with vertebrates in any area of animal physiology, including but not limited to evolutionary and comparative physiology, endocrinology, immunology, and neurophysiology. The successful candidate should be able to teach comparative animal physiology. Second, we seek a biologist whose research uses experimental approaches in **metabolomics** to address fundamental biological questions. Third, we seek a **disease ecologist**, broadly defined, who is interested in understanding the ecology, physiology, and evolution of disease emergence and transmission. Research areas could include epidemiological modeling and pathogen transmission, ecological immunology, genetics and evolution of virulence, importance of host stress on disease dynamics, microbiomes, and others.

UNR offers a highly collaborative research environment with excellent core facilities in proteomics, genomics, flow cytometry, single cell molecular expression, microscopy, bioinformatics, and high-performance computing. A shared instrumentation laboratory offers extensive instrumentation for structure elucidation, including NMR, GC-MS and LC-MS. For all positions an important criterion for selection will be the candidate's ability to enhance one or more existing research strengths in the department.

The department is rapidly growing with five new faculty recruited within the last two years in the areas of neuroscience, evolutionary and functional genomics, and bioinformatics. Currently the department is home to 23 tenure-track faculty, 7 lecturers, and a number of extramurally funded research faculty. Extramural funding is approximately $4,000,000/yr. The faculty participate in interdisciplinary (multi-departmental) graduate programs in molecular and cellular biology (CMB), ecology, evolution and conservation biology (EECB), and a new neuroscience graduate program. See http://www.unr.edu/biology for a listing of the department's research clusters. The faculty also has close ties to the University of Nevada School of Medicine and over $60 million of NIH funds have recently been targeted for biomedical research development on campus. Reno is located in the Sierra Nevada mountains near Lake Tahoe (3 hours from the San Francisco Bay Area), and has been rated one of the best small cities in the US for outdoor recreation and overall quality of life.

To submit application materials, including an application letter, CV, research plans, teaching interests, and contact information for three references go to: **https://www.unrsearch.com/postings/16412 for Organismal Animal Physiology**, **https://www.unrsearch.com/postings/16414 for Metabolomics**, and **https://www.unrsearch.com/postings/16426 for Disease Ecology**. Applications received by **January 12, 2015** will receive full consideration.

EEO/AA Women, under-represented groups, individuals with disabilities, and veterans are encouraged to apply.



AAAS is here –
promoting universal science literacy.

In 1985, AAAS founded Project 2061 with the goal of helping all Americans become literate in science, matematics, and technology. With its landmark publications *Science for All Americans* and *Benchmarks for Science Literacy*, Project 2061 set out recommendations for what all students should know and be able to do in science, mathematics, and technology by the time they graduate from high school.

As a AAAS member, your dues help support Project 2061. If you are not yet a member, join us. Together we can make a difference.

To learn more, visit
aaas.org/plusyou/project2061

AAAS + U = Δ

Cold Spring Harbor Laboratory
2015 Meetings & Courses

Meetings

Systems Biology: Global Regulation of Gene Expression
January 28 - February 1 abstracts due November 18

Cellular Dynamics & Models
March 3 - March 6 abstracts due January 9

Exercise Science & Health
March 9 - March 12 abstracts due January 9

Systems Biology: Networks
March 17 - March 21 abstracts due January 16

Wiring the Brain
March 24 - March 28 abstracts due January 23

Patenting in the Life Sciences: The Patentability of Self-Replicating Systems
March 30 - April 2

RNA & Oliogonucleotide Therapeutics
April 8 - April 11 abstracts due February 6

Fundamental Immunology and Its Therapeutic Potential
April 14 - April 18 abstracts due January 23

The Ubiquitin Family
April 21 - April 25 abstracts due January 30

Telomeres & Telomerase
April 28 - May 2 abstracts due February 6

Biology and Genomics of Social Insects
May 2 - May 5 abstracts due February 13

The Biology of Genomes
May 5 - May 9 abstracts due February 13

The Biology of Cancer: Microenvironment, Metastasis & Therapeutics
May 12 - May 16 abstracts due February 20

Retroviruses
May 18 - May 23 abstracts due February 27

80th Symposium: 21st Century Genetics - Genes at Work
May 26 - May 31 abstracts due March 6

The Evolution of Sequencing Technology: A Half-Century of Progress
July 16 - July 19

Metabolic Signaling and Disease: From Cell to Organism
August 11 - August 15 abstracts due May 22

Eukaryotic mRNA Processing
August 18 - August 22 abstracts due June 2

Mechanisms of Eukaryotic Transcription
August 25 - August 29 abstracts due June 5

Eukaryotic DNA Replication & Genome Maintenance
September 1 - September 5 abstracts due June 12

Microbial Pathogenesis and Host Response
September 8 - September 12 abstracts due June 19

Cell Death
September 15 - September 19 abstracts due June 26

Genome Engineering: The CRISPR/Cas Revolution
September 24 - September 27 abstracts due July 3

Neurobiology of *Drosophila*
September 29 - October 3 abstracts due July 10

Stem Cell Biology
October 7 - October 11 abstracts due July 17

Probabilistic Modeling in Genomics
October 14 - October 17 abstracts due July 24

Genome Informatics
October 28 - October 31 abstracts due August 14

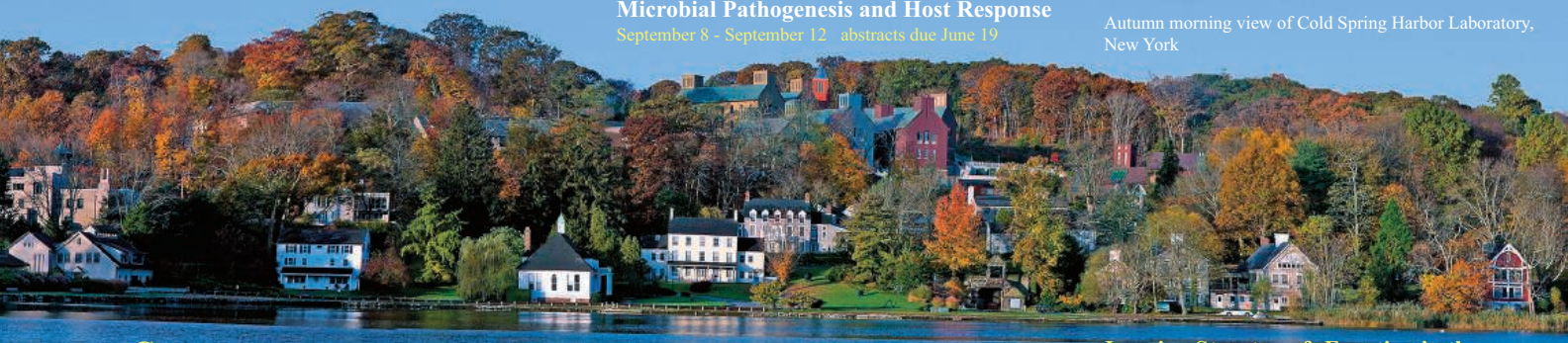
Cell Biology of Yeasts
November 3 - November 7 abstracts due August 21

Single Cell Analyses
November 11 - November 14 abstracts due August 28

Behavior & Neurogenetics of Nonhuman Primates
November 17 - November 20 abstracts due September 4

Plant Genomes & Biotechnology: From Genes to Networks
December 2 - December 5 abstracts due September 18

Rat Genomics & Models
December 9 - December 12 abstracts due September 25



Autumn morning view of Cold Spring Harbor Laboratory, New York

Courses

Workshop on Leadership in Bioscience
March 13 - March 16

Protein Purification & Characterization
April 8 - April 21

Quantitative Imaging: From Cells to Molecules
April 8 - April 21

Cell & Developmental Biology of *Xenopus*: Gene Discovery & Disease
April 9 - April 21

Single Cell Analysis
June 3 - June 16

Advanced Bacterial Genetics
June 3 - June 23

Ion Channels & Synaptic Transmission
June 3 - June 23

Mouse Development, Stem Cells & Cancer
June 3 - June 23

Workshop on Autism Spectrum Disorders
June 4 - June 10

Statistical Methods for Functional Genomics
June 18 - July 1

Workshop on Pancreatic Cancer
June 24 - June 30

***Drosophila* Neurobiology: Genes, Circuits & Behavior**
June 26 - July 16

Frontiers & Techniques in Plant Science
June 26 - July 16

Advanced Techniques in Molecular Neuroscience
June 30 - July 16

Vision: A Platform for Linking Circuits, Perception and Behavior
July 7 - July 20

Proteomics
July 14 - July 27

Eukaryotic Gene Expression
July 21 - August 10

Yeast Genetics & Genomics
July 21 - August 10

Imaging Structure & Function in the Nervous System
July 22 - August 10

Synthetic Biology
July 27 - August 10

Cellular Biology of Addiction
August 4 - August 10

Programming for Biology
October 12 - October 27

X-Ray Methods in Structural Biology
October 12 - October 27

Computational & Comparative Genomics
October 28 - November 3

Antibody Engineering & Phage Display
November 9 - November 22

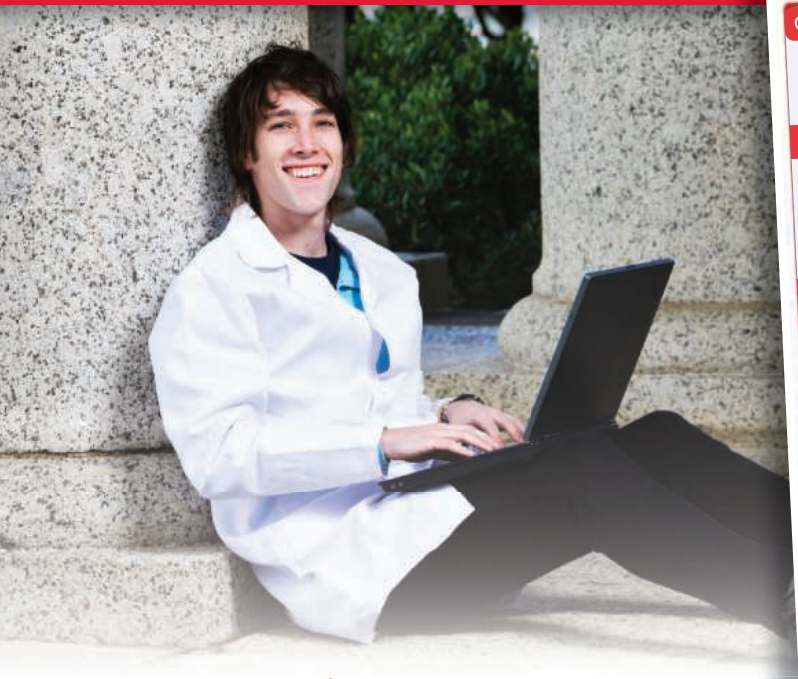
Advanced Sequencing Technologies & Applications
November 10 - November 22

The Genome Access Course
March 30 - April 1 November 16 - November 18

www.cshl.edu/meetings

For your career in science, there's only one Science

A career plan customized *for* you, *by* you.



myIDP.sciencecareers.org



Recommended by leading professional societies and endorsed by the National Institutes of Health, an individual development plan will help you prepare for a successful and satisfying scientific career.

my IDP Science Careers

In collaboration with FASEB, UCSF, and the Medical College of Wisconsin and with support from the Burroughs Wellcome Fund, AAAS and *Science* Careers present the first and only online app that helps scientists prepare their very own individual development plan.

Visit the website and start planning today!
myIDP.sciencecareers.org

In partnership with:

advancing the life sciences 1912-2012 FASEB Federation of American Societies for Experimental Biology

UCSF University of California San Francisco

Medical College of Wisconsin

Burroughs Wellcome Fund

Assistant or Associate Professor
Center for Experimental Therapeutics and Reperfusion Injury
Brigham and Women's Hospital
Harvard Medical School

This Center invites applications for a tenure-track Assistant or Associate Professor position. A PhD and/or MD will join this Center's research mission on mechanisms in the resolution of inflammation and tissue injury. Successful candidate will have maintained research in receptors (GPCR) in inflammation-resolution, with a focus on lipid mediators and resolution. Continued research activities and mentoring are integral to this position. Qualified individuals with post-doctoral training and/or faculty experience should submit a cover letter, and curriculum vitae online at website: **https://research.bwhanesthesia.org/research-groups/cetri/serhan-lab/position**. The names of three references and application should be addressed to **Dr. Charles N Serhan, Search Committee Chair**, no later than **February 9, 2015.**

Successful candidates will maintain independent research in resolution mechanisms in inflammation and tissue injury.

Harvard Medical School is an Equal Opportunity/Affirmative Action Employer.

University of New Hampshire

Assistant Professor in Biological Sciences (4 Positions)

The University of New Hampshire, College of Life Sciences and Agriculture seeks to hire four tenure-track Assistant Professors with demonstrated interests and expertise in: Neurobiology (2), Pharmacology (1) and Bioinformatics (1). Candidates must have a Ph.D. and have demonstrated potential to develop and lead productive research programs. We seek individuals who can integrate their areas of strength with existing academic and research programs. They will be expected to compete successfully in national funding initiatives and to achieve national and international recognition in their fields. Successful candidates will also be expected to train graduate students and teach courses at the undergraduate and graduate levels.

Complete application information is available at: **colsa.unh.edu/employment**. Review of applications will begin on January 20, 2015 and will continue until the positions are filled. Please direct all inquiries to Jane Garnett, COLSA search coordinator, 603-862-5417, **Jane.Garnett@unh.edu**.

All applicants will be required to apply online at the following web site: **https://jobs.usnh.edu.**

UNH is an AA/EEO Employer. UNH is committed to excellence through the diversity of its faculty and staff and encourages women and minorities to apply.

Imperial College London

Department of Epidemiology and Public Health

MRC Early Career Research Fellow

Salary Range: £33,410 to £42,380 per annum

The MRC-PHE Centre for Environment and Health is an internationally recognised centre of excellence for leading pioneering and multidisciplinary research.

The Early Career Research Fellowship programme is a new initiative designed to attract into the Centre exceptional candidates with 0-4 years of postdoctoral experience in the discipline of the proposed fellowship.

The programme will provide a unique opportunity for early-career scientists working in the broad area of environment and health to develop their own competitive, independent research programmes in a supportive, multi-disciplinary environment. The fellowship programme is designed as a platform to apply for subsequent intermediate fellowship applications and similar career progressing schemes.

We intend to recruit, train and develop the next generation of future research and policy leaders from these key and related disciplines:

- Environmental Epidemiology
- Molecular Epidemiology
- Exposure Sciences
- Biostatistics, Computational Biology, Data Science, and Bioinformatics
- Small Area Health Statistics, Spatial Epidemiology, and GIS
- Analytical Sciences (genomics, proteomics, metabolomics, adductomics)
- Mechanistic Toxicology
- Physical, Environmental and Mathematical Sciences

There are no teaching and administrative obligations, and the appointee will be provided with reasonable funds to cover research costs.

Proposals will be judged on the basis of scientific merit, track record of the applicant and potential for development of an independent research programme in the broad area of environment and health. A rigorous and thorough selection procedure akin to the MRC Strategic Fellowship schemes will be adopted.

Applicants should hold a PhD or equivalent in one of the key areas listed above.

This post is full time for a fixed term of 36 months (from 1 April 2015 to 31 March 2018).

For more information and application forms, visit our website http://www.environment-health.ac.uk/centres-fellowship-programme

The fellowship programme will be based at either Imperial College London and/or King's College London. For further information and an informal discussion contact Mr Chun Siew (c.siew@imperial.ac.uk). For formal discussion related to Imperial College London please contact the Centre's Director, Professor Paul Elliott (p.elliott@imperial.ac.uk) or Deputy Director, Professor Frank Kelly (frank.kelly@kcl.ac.uk) for King's College London.

Our preferred method of application is online via our website http://www3.imperial.ac.uk/employment (for employment at Imperial College London) or http://www.kcl.ac.uk/hr/jobs (for employment at King's College London). **Please quote reference number SM265-14**.

Please complete and upload an application form as directed.

Closing date: 22nd February 2015.
Interviews will be held 16th and 17th March 2015.

Imperial Managers lead by example (http://bit.ly/163Lbh7).

Committed to equality and valuing diversity. We are also an Athena Silver SWAN Award winner, a Stonewall Diversity Champion and a Two Ticks Employer and are working in partnership with GIRES to promote respect for trans people.

Public Health England

MRC Medical Research Council

KING'S College LONDON

ScienceCareers online @sciencecareers.org

POSITIONS OPEN

University of Nevada, Reno

The University of Nevada, Reno, a national Tier 1 rank research university, is currently undergoing rapid expansion. As integral components of UNR, the Nevada Agricultural Experiment Station and the College of Agriculture Biotechnology & Natural Resources seek to fill three tenure-track, nine-month positions: (1) **Quantitative Wildlife Population Ecologist** in the Department Natural Resources & Environmental Science; (2) **Insect Disease Transmission** in the Department Biochemistry & Molecular Biology; and (3) **Food Safety** in the Department Agriculture Nutrition & Veterinary Science and in the University of Nevada Cooperative Extension. Successful candidates will teach courses within their Department and establish a rigorous, innovative, and nationally recognized research program in their area of expertise. For more information and to apply, please visit the web page for each position.

Wildlife: https://www.unrsearch.com/postings/16541
Insect: https://www.unrsearch.com/postings/16658
Food: https://www.unrsearch.com/postings/16648

SENIOR FACULTY POSITION
The Johns Hopkins University
School of Medicine
Department of Radiation Oncology and Molecular Radiation Sciences

We seek outstanding individuals with a creative, strong research program focused on DNA damage response, repair, or modulation of these events through chromatin organization. Translational impact of the research focus is highly desirable. Prospective candidates must possess a Ph.D. or M.D.-Ph.D. degree and have an outstanding track record of original research in molecular cancer biology, and successful extramural funding. Laboratory space is located at the Johns Hopkins East Medical Campus in the David H. Koch Cancer Research Building and offers unparalleled resources and opportunities to interact with basic scientists as well as clinical and translational researchers. The position includes a substantial startup package. The successful candidate will be expected to maintain an externally funded research program.

To apply, submit curriculum vitae, a summary of current and future research interests and expected availability date, a description of past research experience and accomplishments, and contact information for three references as a single PDF file as an e-mail attachment to: **Ms. Jamie Janczuk** (**e-mail: jstanko2@jhmi.edu**). Inquires should be sent to: **Director Marikki Laiho, M.D. Ph.D., Division of Molecular Radiation Sciences, Department of Radiation Oncology, The Johns Hopkins School of Medicine, 1550 Orleans Street, Baltimore, MD 21231, e-mail: mlaiho1@jhmi.edu.**

Applications will be assessed on ongoing basis and the deadline for submission is January 15, 2015.

Johns Hopkins University School of Medicine is an Affirmative Action/Equal Opportunity Employer that embraces diversity.

Stop searching for a job; start your career.

ScienceCareers
FROM THE JOURNAL SCIENCE AAAS

ScienceCareers.org

Download your free copy today.

ScienceCareers.org/booklets



From technology specialists to patent attorneys to policy advisers, learn more about the types of careers that scientists can pursue and the skills needed in order to succeed in nonresearch careers.

ScienceCareers
FROM THE JOURNAL SCIENCE AAAS



AAAS is here –

helping scientists achieve career success.

Every month, over 400,000 students and scientists visit ScienceCcareers.org in search of the information, advice, and opportunities they need to take the next step in their careers.

A complete career resource, free to the public, *Science* Careers offers a suite of tools and services developed specifically for scientists. With hundreds of career development articles, webinars and downloadable booklets filled with practical advice, a community forum providing answers to career questions, and thousands of job listings in academia, government, and industry, *Science* Careers has helped countless individuals prepare themselves for successful careers.

As a AAAS member, your dues help AAAS make this service freely available to the scientific community. If you're not a member, join us. Together we can make a difference.

To learn more, visit
aaas.org/plusyou/sciencecareers

AAAS + U = Δ

u b

b
UNIVERSITÄT
BERN

Faculty Position in Biochemistry, University of Bern, Switzerland, with focus on "Membrane transporters: From basic science to drug discovery"

The Medical Faculty of the University of Bern is accepting applications for a faculty position at the level of tenure-track Assistant Professor or Full Professor. The successful candidate will participate in the National Center of Competence for Research (NCCR)[1)] called "TransCure – From Transport Physiology to Identification of Therapeutic Targets" (see: http://www.transcure.org/).

We seek candidates with leadership in basic and applied science related to membrane transporters and ion channels. Applicants must hold an MD, MD/PhD or PhD degree. A medical background would be advantageous. The successful candidate will establish an independent and internationally competitive research program that attracts external funding. In the framework of the TransCure network, he/she will be involved in projects investigating the basic biology, structure-function relationship, biophysics and pharmacology of medically important membrane transporters or channels. He/she will focus on their dysfunction in human diseases as well as pharmaceutical and translational medical perspectives. Preference will be given to candidates who have a documented record of independent research, with publications in high impact journals. Innovative and entrepreneurial thinking as well as demonstrated team building skills and networking abilities will be an asset. He/she will participate in teaching activities in Biochemistry, Molecular Biology and Pharmacology for students in Medicine, Veterinary Medicine and Pharmacy. For teaching, fluency in German and English is a prerequisite.

The University of Berne is an equal opportunity employer. Women are particularly encouraged to apply in accordance with the NCCR's effort to increase the number of female scientists.

For more information, please contact the Chairman of the Search Committee, Prof. Ernst Niggli, Department of Physiology, CH-3010 Bern, Switzerland (niggli@pyl.unibe.ch).

Please send applications together with CV, list of publications, list of funding, a short outline of planned research and names of three references electronically to the Office of the Dean (bewerbungen@meddek.unibe.ch). Guidelines for applications can be found on: http://www.medizin.unibe.ch/content/akademische_laufbahn/vorgehen_bewerbungen/.

The submission deadline is 19 January 2015.

[1)] NCCR is a research instrument of the Swiss National Science Foundation; NCCR TransCure

The Dean's Office, Faculty of Medicine, University of Bern, Murtenstrasse 11, CH-3010 Bern

By **Jessica K. Polka** *and* **Kristin A. Krukenberg**

Making science a desirable career

Most scientists (ourselves included) headed off to graduate school eager to discover new things, test crazy ideas, and hopefully change the world, or at least explain a piece of it. Sadly, these days such aspirations are overshadowed by fierce competition for papers, grants, and the tenured positions trainees need to build independent careers in academic research. Even those who have attained one of those rare, coveted professorships are under intense strain: A preoccupation with funding and other career-related metrics crowds out real discovery among aspiring and established scientists alike. The result: Science has become a less desirable career. We can fix the problem, but only if we're willing to make fundamental changes to how academic science is structured and funded.

For decades, trainees—graduate students and postdocs—have performed most of the day-to-day work of academic science. Motivated by the promise of satisfying careers, they are willing to work long and hard for minimal pay. For science, that works out well in the short term but not the long: The distinction between trainee and employee blurs. When more is invested in science—a good thing—the ranks of graduate students and postdocs swell to meet the demand for temporary labor. Permanent positions don't keep pace, eroding the economic and intellectual value of scientific training. Trainees take ever longer—into their mid-30s and beyond—to accrue credentials sufficient to stand out and get hired for serious, long-term faculty jobs. With each passing year, scientists young and old spend more time worrying about their careers and less time making discoveries. As long as academic science is structured the way it is, we don't really have a choice.



"Career instability ... turns some of the best talent away."

Some argue that intense competition strengthens science by allowing only the fittest to survive. But the incentives that arise from such competition run counter to creativity, openness, and innovation. We may even be selecting for undesirable traits, such as recklessness and a focus on self-marketing over scientific substance. The weight put on publishing papers in high-impact journals may encourage scientists to exaggerate their findings, or worse. In short, it is evident that hypercompetition is inconsistent with "affording the prepared mind complete freedom for the exercise of initiative," which Vannevar Bush, who devised our system of academic research funding in the 1940s, considered to be of "supreme importance."

So why do students even start down the road to science careers? We believe it's a combination of ignorance of the realities, overconfidence, and a punch-drunk passion for science. At some point during training, however, that enthusiasm starts to wear off. Every trainee has encountered exceptionally talented students who would have made major contributions but decided to leave science—its uncertain prospects, risk-averse funding, and incentives for distortion and dishonesty—in favor of more practical and lucrative careers.

To draw the best young minds to research and then keep them here, we need to change how we train scientists and how academic science gets done. More of the work should be done by staff scientists in stable positions and less of it by (fewer) trainees. We should move the career bottleneck from faculty to the postdoc or even to graduate school admissions. Training grants and fellowships should replace grant support for most trainees so that more can be invested in research grants without increasing the number of trainees and further glutting Ph.D.-level science labor markets. Such measures would have useful side effects: encouraging better training and intellectual independence. More important, they would improve young scientists' career prospects.

It is time to address the career instability that turns some of the best talent away from science. The effort and investment needed to make research careers more attractive will be handsomely repaid in the form of future discoveries. ■

Postdocs Jessica K. Polka and Kristin A. Krukenberg are organizers of the Future of Research Symposium held in Boston in early October and corresponding authors for the resulting report. For more on life and careers, visit www.sciencecareers.org.

ILLUSTRATION: ROBERT NEUBECKER

Downloaded from www.sciencemag.org on December 14, 2014

Published by AAAS



10-fold higher transfection efficiency in difficult cells

> We were very happy and surprised to see Lipofectamine® 3000 reagent provide a more than 10-fold difference in transfection efficiency in our difficult-to-transfect cell line. There was even a reduction in cell death. Awesome results!
>
> **—Rui Eduardo Castro, PhD, University of Lisbon**

Discover the versatile reagent with the power to transfect your most difficult cells.

Find out more at **lifetechnologies.com/3000**

Join the conversation at #DiffCells

life technologies

A **Thermo Fisher Scientific** Brand

For Research Use Only. Not for use in diagnostic procedures. © 2014 Thermo Fisher Scientific Inc. All rights reserved.
All trademarks are the property of Thermo Fisher Scientific and its subsidiaries unless otherwise specified. **CO012640 0714**



This is the start of something big.

Science Advances | AAAS

SIGNIFICANT RESEARCH, GLOBAL IMPACT

Introducing *Science Advances* – the new, online-only, open-access journal from *Science* and AAAS. Find out how you can be among the first authors published at **scienceadvances.org.**

NEW ENGLAND BioLabs® Inc.

PASSION IN SCIENCE AWARDS™

CELEBRATING 40 YEARS OF PASSION IN SCIENCE

New England Biolabs is proud to be recognized as a world leader in the discovery, production and supply of reagents for the life sciences. For 40 years, NEB has embraced the core values of scientific pursuit, environmental stewardship, humanitarian efforts and artistic expression.

In celebration of our 40th anniversary, we recently introduced the Passion in Science Awards™, recognizing those within the scientific community working to solve many of today's challenges.

Based on NEB's foundational values, the Passion in Science Awards are offered in four categories:

- Inspiration in Science
- Humanitarian Duty
- Environmental Stewardship
- Arts and Creativity

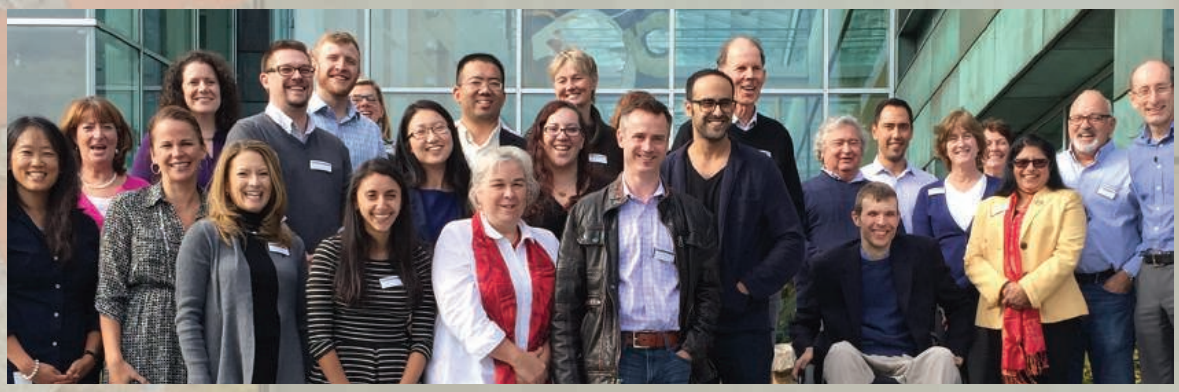
This October, fifteen dedicated and compassionate scientists were chosen from around the world to participate in this first-of-its-kind event. Award winners visited NEB's campus in Ipswich, MA, for two days of discussions, presentations, tours and an awards celebration.

Visit www.neb.com/PassionInScience to learn more about our award winners and their inspiring stories.

NEW ENGLAND BIOLABS® and NEB® are registered trademarks of New England Biolabs, Inc.
PASSION IN SCIENCE AWARDS™ is a trademark of New England Biolabs, Inc.

Artwork shown on side of building created by Nicole Gordon.

2014 PASSION IN SCIENCE AWARDS



The 2014 Passion in Science Award winners are pictured with their hosts outside the New England Biolabs' facility in Ipswich, MA.





Left, Award winner Shelly Xie shapes sand with her hands to convey the hardship of families in developing nations, who battle parasitic infections. Right, winner Tal Danino developed methods to micropattern cervical cancer cells and collaborated with artist Vik Muniz to create art that raises money for research.





Left, Award Winner Peter Hotez answers questions about his work with neglected tropical diseases. Right, Award winners Jason Furrer and Ite Laird-Offringa admire their custom, handcrafted slate award plaques.

Simplified analysis of lipid or detergent content in biological samples using the IR-based Direct Detect® Spectrometer

Lipids are a fundamental class of biologically relevant compounds. Despite the growth in lipid research, analytical methods applied for their characterization typically involve multistep procedures, requiring extensive sample manipulation and separation from other biomolecules, like proteins, before the analysis. Lipids can be isolated by liquid–liquid extraction and separated into classes, often derivatized, and then analyzed. Methods of lipid analysis include measuring iodine value, elemental phosphorus, acid value (saponification equivalent), peroxide value and radiochemical techniques or using chromatography and mass spectrometry.

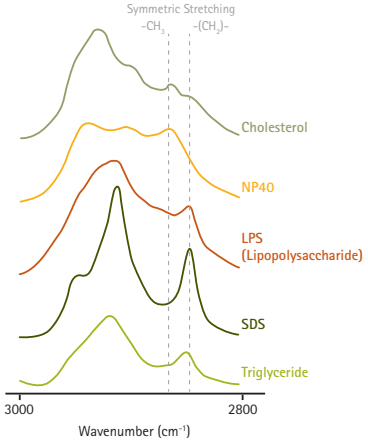


Figure 1.
C-H symmetric stretching bands observed in the infrared spectra of lipids.

The infrared (IR)-based Direct Detect® spectrometer enables simultaneous protein quantitation and lipid analysis, directly from a biological sample. Previous reports have also identified IR spectroscopy as a viable method for lipid analysis[1]. Characteristic lipid bands in IR spectra, such as the aliphatic group stretching (2800 - 3000 cm^{-1}), ester C=O stretching (around 1740 cm^{-1}) or phosphate stretching (around 1235 cm^{-1}) permit qualitative and quantitative analysis of lipid content[2]. Given that each lipid possesses an IR signature uniquely defined by its chemical composition and structure, IR spectroscopy offers a means of qualitative lipid discrimination, as well as quantitation of known lipid(s) in cases where a viable standard curve has been determined.

The Direct Detect® spectrometer utilizes the C-H symmetric stretching vibrational population between 2870 and 2840cm^{-1} to determine lipid or detergent content (Figure 1). The Direct Detect® assay-free sample card enables analysis of aqueous-based biological samples, which are normally not compatible with infrared spectroscopy, due to their high water content. These assay-free cards are also compatible with many organic solvents.

Materials and Methods

Measurements of sample concentration were acquired using Direct Detect® assay-free sample cards (EMD Millipore, Cat. No. DDAC00010-8P) and the Direct Detect® spectrometer (EMD Millipore, Cat. No. DDHW00010-WW). Each assay-free card contains four polytetrafluoroethylene (PTFE) membrane positions, sized for easy sample application and analysis. All measurements were performed using 2 µL of sample solution per membrane position. Unknown lipid mixtures were analyzed in the "Relative Absorbance" mode, where the system delivers information based solely on IR signal strength. Empirical sample concentration values were determined by interpolation from calibration curves developed for each specific lipid or detergent. For the experiments reported here, the system was calibrated using Tetracosanoic Acid in chloroform and 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate (CHAPS) in phosphate-buffered saline (PBS). A series of seven concentrations from 0.25 mg/mL to 1.75 mg/mL was used to generate an instrument calibration curve for Tetracosanoic Acid. For CHAPS, the calibration curve was formed using a series of seven concentration points spanning 0.25% to 4%. The strength of IR signal for each concentration was fitted to a regression line represented by linear equations y = 0.04519x - 0.00679 (lipid) and y = 0.01774x + 0.00442 (detergent). These equations were used by the Direct Detect® software to determine the concentration of Tetracosanoic Acid and CHAPS in subsequent samples.

Robustness of the Relative Absorbance mode was further demonstrated through simultaneous analysis of protein and lipid content in samples containing mixture of both biomolecules as well as in breast cancer tissue lysate. Frozen tissue, derived from a breast ductal carcinoma, was divided into 2 equal samples. Tissue was covered with 2 mL RIPA buffer (EMD Millipore, Cat. No. 20-188) or CytoBuster™ protein extraction reagent (EMD Millipore, Cat. No. 71009-50mL), both supplemented with an inhibitor cocktail, and disrupted with a glass tissue homogenizer. Effective removal of the fatty fraction from the resulting tissue homogenate by a series of centrifugation steps was also monitored by the Direct Detect® spectrometer.

Results

Accuracy of lipid and detergent quantitation

The accuracy of concentration estimation within the dynamic ranges established for Tetracosanoic Acid and CHAPS was assessed using 0.8 mg/mL Tetracosanoic Acid in chloroform and 1.8 % CHAPS in PBS. Within a well-defined calibration method, the instrument was capable of estimating lipid and detergent concentration with minimal error: the Direct Detect® spectrometer estimated the concentration of the Tetracosanoic Acid samples to be 0.853 ± 0.14 mg/mL (2.4%CV) and the CHAPS samples were estimated to be 1.8 ± 0.004% (2.3%CV).

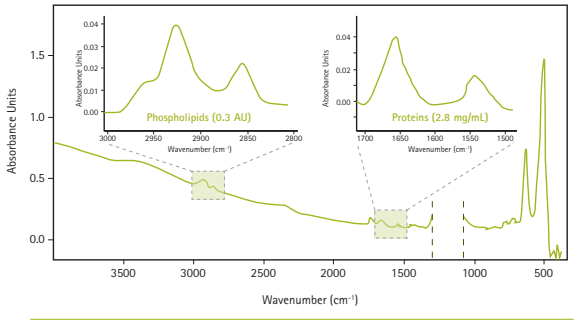


Figure 2. Analysis of protein and lipid content from a single sample measurement. Because the sample composition and ratio of individual components were unknown, the analysis of the phospholipids was performed using the "Relative Absorbance" mode.

EMD MILLIPORE

Protein quantitation and lipid content analysis performed using the same measurement

IR spectral profiles of proteins and lipids are distinct, thereby enabling the simultaneous quantitation of proteins and analysis of lipids using the same measurement. Protein content can be quantified using one of the pre-loaded standard curves (all prepared in PBS pH 7.4)[3]. Lipid analysis can be performed, using the same spectra, in the "Relative Absorbance" mode. We determined phospholipid concentration values in a complex mixture based on the signal strength of the IR profile for the C-H symmetric stretching region (Figure 2). The protein in the sample was quantified (using NIST BSA AM1 calibration method) at 2.8 mg/mL.

Monitoring the lipid profile during the preparation of a breast cancer tissue lysate

In biological research, downstream detection and analysis methods are often strongly influenced by the mode of lysate preparation. Traditionally, quantitation of proteins and lipids in lysates has been tedious, requiring large sample volumes and specialized methodologies, and the results for either component obtained using classical methods may be obscured by cross-interference. The Direct Detect® spectrometer has enabled rapid analysis of total protein with simultaneous monitoring of lipid content, thereby simplifying and improving the analytical process[4]. Breast cancer tissue lysates were prepared using RIPA buffer and CytoBuster™ protein extraction reagent. Following tissue homogenization, the Direct Detect® spectrometer was used to monitor the efficiency of fat removal and total protein concentration during centrifugal extraction. Direct Detect® spectrometer data showed gradual removal of a fatty fraction from the samples (Figure 3). The same spectra were also used to determine the total protein recovery across the various fractions (Table 1). From this limited study, it is clear that the Direct Detect® system offers a means for in-line process optimization for maximal yield and/or purity.

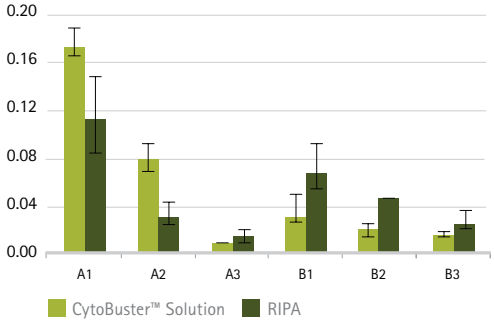


Figure 3. Lipid content as determined using the Direct Detect® spectrometer in the "Relative Absorbance" mode in 3 different fractions of breast cancer tissue lysates prepared with CytoBuster™ Protein Extraction Reagent or RIPA buffer.

Sample	Spin condition & fraction collected	Total protein content (mg/mL) CytoBuster™ Protein Extraction Reagent	RIPA
A1	spin @10,000 xg (top fatty fraction)	5.0	14.0
A2	spin @15,000 xg (top fatty fraction)	2.7	20.0
A3	spin @15,000 xg (bottom layer)	5.0	17.0
B1	spin @10,000 xg (top fatty fraction)	3.3	3.6
B2	spin @15,000 xg (top fatty fraction)	2.1	5.0
B3	spin @15,000 xg (bottom layer)	2.8	5.0

Table 1. Total protein recovery from breast cancer tissue lysed with RIPA buffer or CytoBuster™ protein extraction reagent measured using the Direct Detect® spectrometer.

Conclusions

The Direct Detect® spectrometer enables rapid analysis of lipids and detergents in addition to accurate and reproducible protein quantitation. The ability to simultaneously monitor protein concentration and fat removal during sample preparation provides a tool for assay optimization as well as greater confidence in final sample purity.

To learn more about Direct Detect® spectrometer, visit

www.merckmillipore.com/DirectDetect

References

1. Pidgeon C., Apostol G. and Markovich R. Anal. Biochem. 1989, 181(1): 28-32.
2. Pistorius A.M.A, DeGrip W.J. and Egorova-Zachernyuk T.A. Biotechnol Bioeng. 2009, 103(1): 123-9.
3. EMD Millipore. Literature No. AB3355ENEU, 2012.
4. Gutierrez, S. et al. Poster Presentation, Human Proteome Organization (HUPO) 11th Annual World Congress, Boston, MA, USA, September 2012.

Goodbye, Bradford Assays!
Drive your research forward with IR-based quantitation.

With Direct Detect® IR-based spectrometer, you can distinguish proteins and peptides from sample components, such as lipids, that interfere with classical quantitation methods. Now you can achieve truly accurate results without the pitfalls of colorimetric assays, even for most lysates and complex samples.

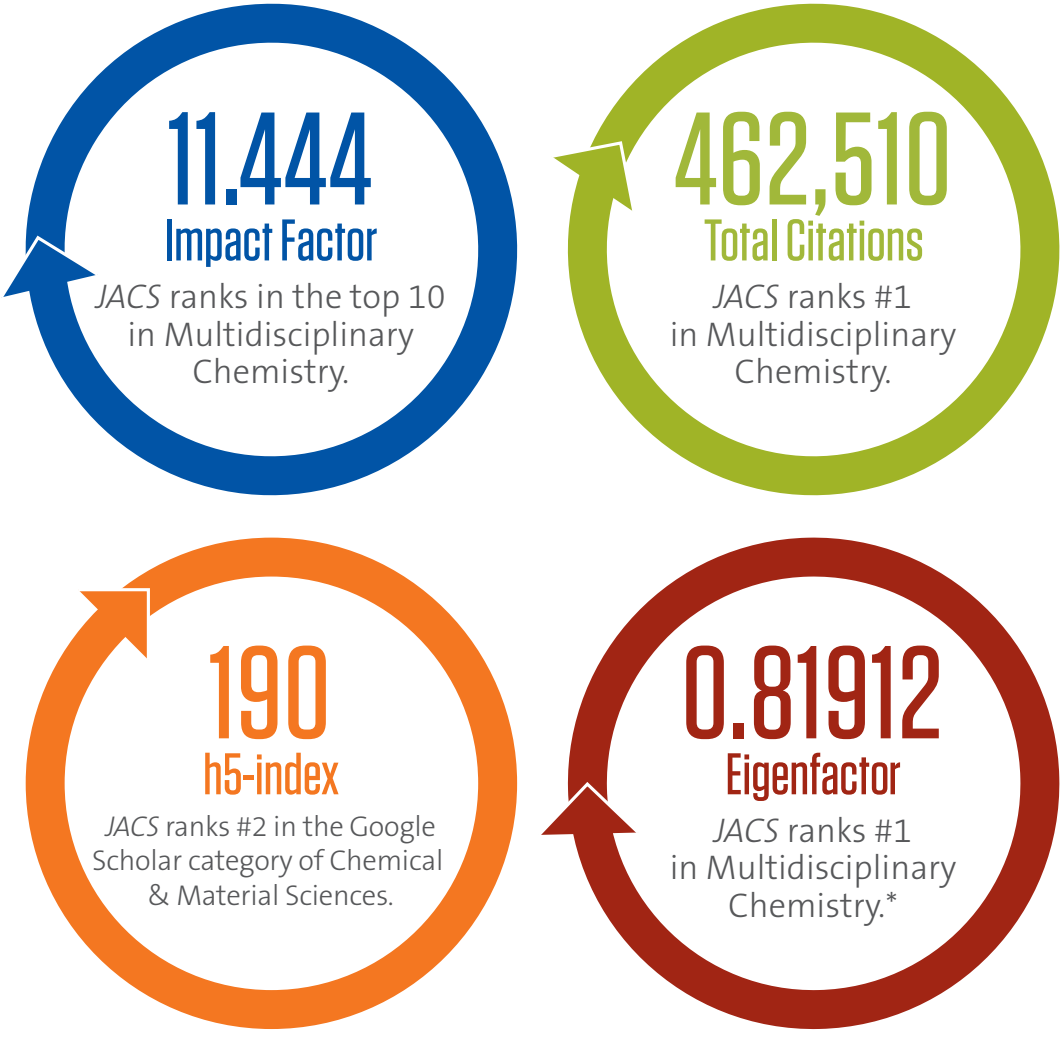


Quit Assays Forever—Quantitate Directly.
Learn more at: www.emdmillipore.com/DirectDetect

EMD Millipore is a division of Merck KGaA, Darmstadt, Germany

EMD Millipore, the M logo and Direct Detect are registered trademarks of Merck KGaA, Darmstadt, Germany. CytoBuster is a trademark of Merck KGaA, Darmstadt, Germany. © 2014 EMD Millipore Corporation, Billerica, MA USA. All rights reserved. BS-GEN-14-10790.

There are many ways to measure a journal...



Here's another way: Look at its editors. »»

J|A|C|S
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

ACS Publications
Most Trusted. Most Cited. Most Read.

As reported in the 2013 Journal Citations Reports® from Thomson Reuters.

*What is Eigenfactor? Go to **www.eigenfactor.org**.

JACS editors: distinguished

Unlike journals with editors who are full-time publishing professionals, *JACS* offers you the benefit of having your manuscript submission handled by scientists who are engaged in cutting-edge research and still publishing their own papers. That's why *JACS* editors have a better idea of what is important to the science, where a field is headed, and what is important to you as an author.

At *JACS*, serving the scientific community is what matters most. That's the real measure of our success.



EDITOR

Peter J. Stang
Distinguished Professor
David P. Gardner
Presidential Chair
University of Utah

ASSOCIATE EDITORS



Eric V. Anslyn
University Distinguished Teaching Professor and Norman Hackerman Professorship in Chemistry
The University of Texas at Austin



Weston Thatcher Borden
Distinguished University Research Professor and Welch Chair
University of North Texas



Kara L. Bren
Professor of Chemistry
University of Rochester



Benjamin F. Cravatt
Professor and Chair, Department of Chemical Physiology
The Scripps Research Institute, Scripps California



Paul S. Cremer
J. Lloyd Huck Chair in Natural Sciences
The Pennsylvania State University



Lyndon Emsley
Professor of Chemistry
Université de Lyon, France



Jean M. J. Fréchet
Vice President for Research and Distinguished Professor of Chemistry, King Abdullah University of Science and Technology, Saudi Arabia



Gregory C. Fu
Altair Professor of Chemistry
California Institute of Technology



Miguel A. García-Garibay
Professor and Chair, Department of Chemistry and Biochemistry
University of California, Los Angeles



Martin Gruebele
James R. Eiszner Endowed Chair in Chemistry, Professor of Physics and Center for Biophysics and Computational Biology
University of Illinois at Urbana-Champaign



Sidney M. Hecht
Director, Center for BioEnergetics, Biodesign Institute, and Professor of Chemistry
Arizona State University



Taeghwan Hyeon
Distinguished University Professor and Director, Center for Nanoparticle Research, Institute for Basic Science
Seoul National University, Korea

J|A|C|S
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Follow *JACS* at **facebook.com/JAmChemSoc**

Follow *JACS* on Twitter **@J_A_C_S**

and active researchers



William D. Jones
C. F. Houghton Professor of Chemistry
University of Rochester



Thomas E. Mallouk
Evan Pugh Professor, Associate Director, Penn State MRSEC, Director, Center for Solar Nanomaterials
The Pennsylvania State University



Chad A. Mirkin
Director of the International Institute for Nanotechnology, the George B. Rathmann Prof. of Chemistry, Prof. of Chemical and Biological Engineering, Prof. of Biomedical Engineering, Prof. of Materials Science & Engineering, and Prof. of Medicine
Northwestern University



Klaus Müllen
Professor and Director, Max Planck Institute for Polymer Research, Mainz, Germany



Eiichi Nakamura
Professor of Physical Organic Chemistry
The University of Tokyo, Japan



William R. Roush
Professor of Chemistry, Associate Dean of Graduate Studies, The Scripps Research Institute, Florida



Melanie S. Sanford
Moses Gomberg Collegiate Professor of Chemistry, Arthur F. Thurnau Professor of Chemistry
University of Michigan



Alanna Schepartz
Milton Harris '29 Ph.D. Professor of Chemistry, Professor of Molecular, Cellular, and Developmental Biology
Yale University



Matthew S. Sigman
Presidential Endowed Chair of Chemistry
University of Utah



Donald G. Truhlar
Regents Professor
University of Minnesota



Li-Jun Wan
Professor and Director, CAS Key Laboratory of Molecular Nanostructure and Nanotechnology, Director, Beijing National Laboratory for Molecular Sciences, Institute of Chemistry, Chinese Academy of Sciences



Henry S. White
Distinguished Professor
University of Utah



Karen L. Wooley
W. T. Doherty-Welch Chair and University Distinguished Professor
Professor of Chemistry, Chemical Engineering, and Materials Science & Engineering
Director, TAMU Laboratory for Synthetic-Biologic Interactions
Texas A&M University



Omar M. Yaghi
The James and Neeltje Tretter Professor of Chemistry
University of California, Berkeley



Peidong Yang
S.K. and Angela Chan Distinguished Professor of Energy, Professor of Chemistry
University of California, Berkeley

ACS Publications
Most Trusted. Most Cited. Most Read.

Submit your best research to *JACS* at pubs.acs.org/jacs



PERCEPTIONS

Science and Religious Communities
A National Conference

Join us at the Ronald Reagan Building and International Trade Center in Washington DC, March 13, 2015 for Perceptions: Science and Religious Communities, a one-day national conference that will bring together leaders in science and religion for conversation and relationship building.

Insights by national leaders—including Nobel Prize winning physicist William D. Phillips, climate scientist Katharine Hayhoe, Eugenie C. Scott of the National Center for Science Education, and National Association of Evangelicals president Leith Anderson—will combine with lively topical discussions and casual conversation for an enriching day of dialogue.

DoSER
Dialogue on Science, Ethics, & Religion

AAAS

Register at **www.aaas.org/doser**. More information about the Perceptions Project can be found at **www.perceptionsproject.org**

AAAS *Travels*



Wild Iceland
June 13-22, 2015

Land of Fire & Ice & Nature's Paradise!

Grab your binoculars & cameras & discover Iceland—the world's oldest democracy! $3,995 pp + air

For a detailed brochure, call (800) 252-4910
All prices are per person twin share + air

BETCHART EXPEDITIONS Inc.
17050 Montebello Rd, Cupertino, CA 95014
Email: AAASInfo@betchartexpeditions.com
www.betchartexpeditions.com

Exploring the Frontiers of Science in Western China



China's remarkable economic rise is also propelling advances across the sciences, and these intertwined trends are transforming the western section of the country. To speed up this transformation, the Chinese Academy of Sciences (CAS), universities, and the central government are all creating award schemes aimed at attracting leading scholars trained abroad to take up positions in western China. Science centers across the region are ramping up efforts to make research teams more innovative and more international.

See the full story on page 1387.

Upcoming Features
Faculty—January 30, 2015 ■ *Postdocs—February 27, 2015*

Produced by the *Science*/AAAS Custom Publishing Office

Quantitative Western Blot System

The Amersham WB system is a fully integrated system for quantitative SDS-PAGE and Western blotting of proteins using fluorescence detection. The system is designed to drive down the variability seen with conventional Western blotting. With the Amersham WB system, every stage of the Western blotting process, including electrophoresis, transfer, probing, and scanning, is standardized and monitored—resulting in consistent, quantitative protein analysis. Western blotting requires a complex combination of analytical knowledge, extensive manual handling, dexterity, and strict adherence to protocols in order to achieve reliable results. Failure rates, however, can be as high as 25% and data variations between different users can be as high as 35%–40%. To address this shortfall in consistency, GE Healthcare Life Sciences has developed the Amersham WB system, with a standardized traditional Western blotting workflow that typically achieves a variability of less than 10% between users. The system's consistency and reproducibility delivers robust quantitative data.

GE Healthcare Life Sciences
For info: +44-(0)-0800-032-9201
www.gehealthcare.com

Protein A Kit

The Gyrolab Protein A Kit is designed for use on the Gyrolab xP workstation. During purification of biotherapeutics, Protein A can leach from the chromatography support and co-elute with the therapeutic antibody product. The contamination complicates subsequent analysis using immunoassays. In addition, Protein A can elicit an immune response. It is therefore essential to accurately quantify Protein A during the purification process, and prior to product release. The new Protein A kit, designed and validated for use with Gyros nanoliter-scale immunoassay platform, includes automated acid dissociation of Protein A and IgG during sample preparation, to ensure contamination is accurately assessed. Gyros has incorporated proprietary antibodies to ensure high specificity and sensitivity. The easy-to-use kit quantifies native Protein A, recombinant Protein A variants, and MabSelect SuRe ligand, and has fewer manual steps than other standard enzyme-linked immunosorbent assay kits, reducing hands-on time and risk for error.

Gyros
For info: +46-(0)-18-56-63-00
www.gyros.com



Western Blot Imaging System

Until now, digital imaging systems have failed to deliver the sensitivity and resolution of film. The new ChemiDoc Touch System, however, allows detection of faint bands missed by film and produces publication-quality images. The ChemiDoc Touch outperforms film in other ways as well. When using film to image abundant proteins, strong bands quickly saturate and become unquantifiable. Saturated bands can also obscure the signal from adjacent faint bands, making Western blotting with film challenging. The ChemiDoc Touch addresses these issues through its wide dynamic range, which permits easy and reliable quantification even of highly abundant proteins, and through the instrument's ability to optimize exposure for each protein of interest. Capturing images with the ChemiDoc Touch is easy. The ChemiDoc Touch offers a smooth, intuitive user experience that makes capturing, reviewing, selecting, and exporting images efficient and straightforward. The imager also allows stain-free imaging, a technology exclusive to Bio-Rad.

Bio-Rad Laboratories
For info: 800-424-6723
www.bio-rad.com/chemidoctouchpr

Phosphorylation Detection Assays

AlphaLISA SureFire Ultra assays are designed to reduce assay time and to improve the protein phosphorylation detection process. With high sensitivity and wide dynamic range, AlphaLISA SureFire Ultra assays are quick and easy to run with a minimal amount of sample, significantly reducing time and expense. The AlphaLISA SureFire Ultra assays use TGR BioSciences' CaptSure technology to eliminate antibody interference in the therapeutic antibody screening process. AlphaLISA SureFire Ultra sensitivity and dynamic range also make it ideal for other types of screening, such as small molecule modulation, detection of low and basal levels of phosphorylation, and detection of phosphorylation in tissue lysates. The assays also leverage PerkinElmer's AlphaLISA technology that has better spectral characteristics related to hemoglobin interference, which can otherwise hinder the accuracy of measurements. AlphaLISA SureFire Ultra generates results in less than two and a half hours with only one or two incubation steps and no washing.

PerkinElmer
For info: 800-762-4000
www.perkinelmer.com/surefire

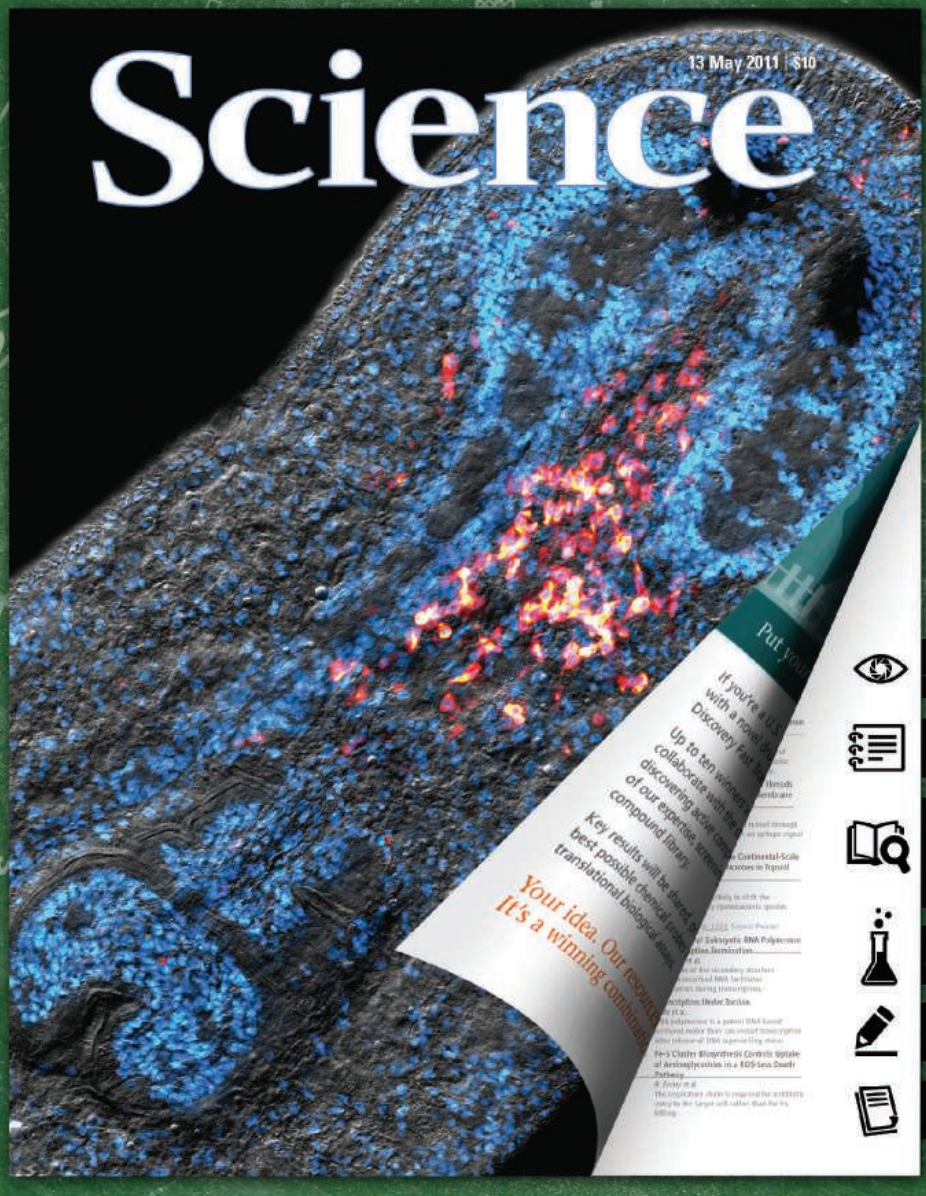
Remote Access Imaging System

The vibrant new T:Genius imaging system allows high-performance, walkaway imaging of gels and blots on a smartphone or tablet. Using the T:Genius' "StatusLink" feature, researchers can stay updated and even share image results with colleagues, no matter whether they are down the hall or in another country. This is a great time saver when, for instance, scientists are developing a long-exposure chemi blot and want to see how it's doing while having lunch or busying themselves with something else in the lab. The new T:Genius is both stylish and practical, comes in a choice of exciting colors, and features a sensitive camera that will give high-res images. With the powerful GeneSys image capture software, researchers can set up their imaging protocols, click, walk away, and see accurate DNA or protein results later in the lab or on their tablet or phone.

Syngene
For info: +44-(0)-1223-727123
www.syngene.co.uk/t-genius

Electronically submit your new product description or product literature information! Go to www.sciencemag.org/products/newproducts.dtl for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.



Learning Lens

Learning Notes

Connect to Learning Standards

Activities

Discussion Questions

References

Online tools to help your students analyze a professional research paper!

Science

Tell me and I forget. Teach me and I remember. Involve me and I learn. -- Benjamin Franklin

Featuring over 6 research papers at any given time, *Science* in the Classroom is specifically designed to help young researchers understand the structure and workings of professional scientific research.

Learn for yourself how *Science* in the Classroom can help your students deepen their understanding of scientific research. **Visit scienceintheclassroom.org today.**

AAAS
ADVANCING SCIENCE, SERVING SOCIETY



Learn more and don't let your job search leave you washed up.

- Search thousands of job postings
- Create job alerts based on your criteria
- Get career advice from our Career Forum experts
- Download career advice articles and webinars
- Complete an individual development plan at "myIDP"

Target your job search using relevant resources on **ScienceCareers.org**.

ScienceCareers
FROM THE JOURNAL SCIENCE AAAS



Have your cake and eat it too
R&D Systems Luminex® Assays

Get the **most plexible analytes** and enjoy
the **accuracy and precision you expect**
from R&D Systems products.

Find out how to win a cake for your lab.
rndsystems.com/LabCake

R&D systems™
a bio-techne brand

Luminex is a registered trademark of Luminex