

The tarnished legacy of
a master mentor p. 814

Restoring sight
to the blind p. 827

Carbon nanotube
circuits pp. 850, 874, & 878

Science

\$15
22 MAY 2020
sciencemag.org

 AAAS

FIGHTING BLIGHT

A wheat fungus meets
its match pp. 822 & 844

CONTENTS

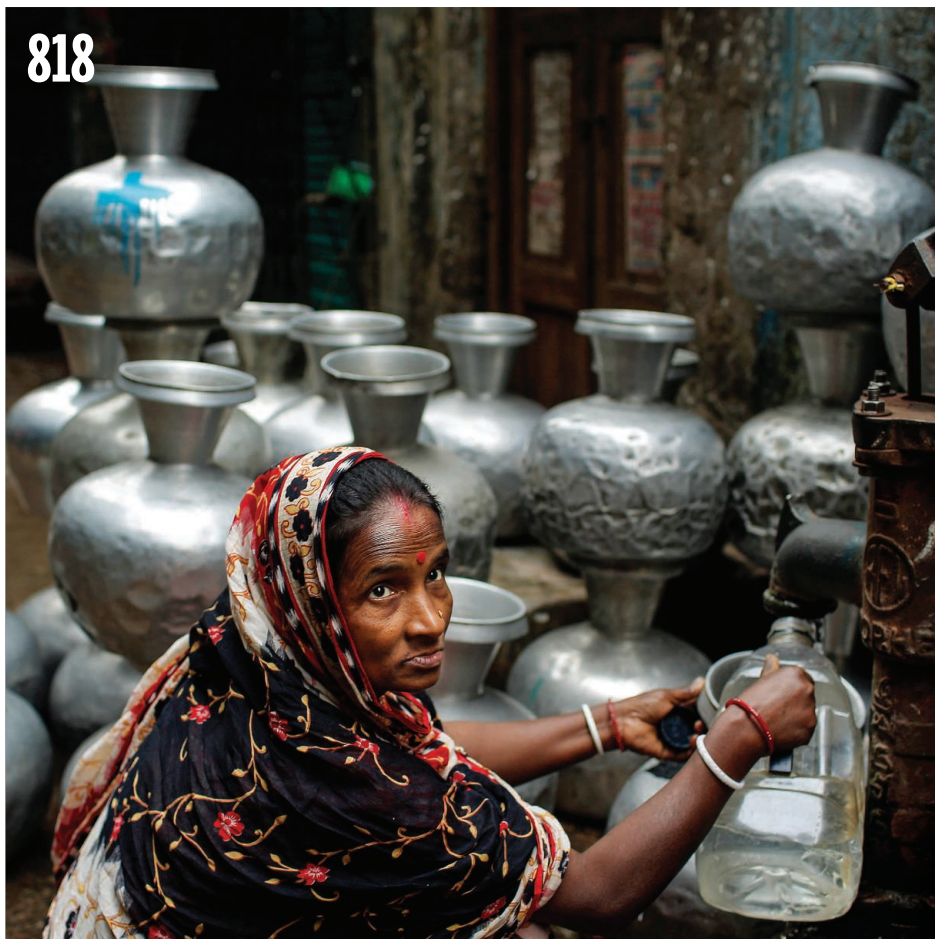
22 MAY 2020 • VOLUME 368 • ISSUE 6493



835

Seeking the
butterfly's secrets

818



NEWS

IN BRIEF

804 News at a glance

IN DEPTH

807 Long-acting drug acts like a short-term AIDS vaccine

Simpler than daily pills, injections to prevent HIV infections protect people for 2 months *By J. Cohen*

808 Case clustering emerges as key pandemic puzzle

Why do some patients infect many others, whereas many don't spread the virus at all?

By K. Kupferschmidt

809 T cells found in coronavirus patients 'bode well' for long-term immunity

New findings suggest past infections may offer some protection against the novel coronavirus *By M. Leslie*

810 AI systems aim to sniff out coronavirus outbreaks

Alerts would not substitute for on-the-ground sleuthing, experts warn *By A. Cho*

811 Crewed launch deepens ties between NASA and SpaceX

Bold company's rise brings benefits and risks for space science *By A. Mann*

813 Tropical forests store carbon despite warming

But if global temperatures reach key threshold, dying trees will release warming gases *By E. Pennisi*
REPORT P. 869

FEATURES

814 Hard lessons

Carlos Castillo-Chavez trained a record number of minority mathematicians—but his tough-love approach took a toll
By J. Mervis and M. Stellino

INSIGHTS

PERSPECTIVES

818 Global solutions to a silent poison

Modeling arsenic in domestic well water highlights large data gaps in testing
By Y. Zheng

REPORT p. 845

820 Tracking both ultrafast electrons and nuclei

Electron diffraction correlates the excited-state decay of pyridine with its ring distortion
By W. Domcke and A. L. Sobolewski

REPORT p. 885

821 Solitons and topological waves

A laser-fabricated waveguide array creates a nonlinear medium that supports solitons
By Mark J. Ablowitz and Justin T. Cole

REPORT p. 856

822 Breeding a fungal gene into wheat

An ancient cross-kingdom gene transfer enables wheat resistance to a fungal toxin
By B. B. H. Wulff and J. D. G. Jones

REPORT p. 844

824 The secret lives of bees as horticulturists?

Pollen-starved bumble bees may manipulate plants to fast-forward flowering
By L. Chittka

REPORT p. 881

825 Synchronized to an optical atomic clock

Microwave generation using optical frequency comb technology hits new milestones
By E. A. Curtis

REPORT p. 889

827 Restoring vision to the blind

Ideas abound to restore vision to people blinded by retinal disease
By J. E. Dowling

829 Rapid repurposing of drugs for COVID-19

The emergence of a new coronaviral respiratory disease calls for repurposing existing drugs *By R. K. Guy et al.*

831 John Horton Conway (1937–2020)

Innovative mathematician and passionate educator *By M. Baker*

POLICY FORUM

832 Ethics of controlled human infection to address COVID-19

High social value is fundamental to justifying these studies *By S. K. Shah et al.*

BOOKS ET AL.

835 Butterflies and the people who love them

A meandering investigation hints at how much is left to learn about these charismatic insects *By A. A. Agrawal*

836 Diversity and the dinner plate

An intimate exploration of foraged flavors brings nuance to the wild food discussion *By L. Newman*

LETTERS

838 China's wild turtles at risk of extinction

By J. Wu et al.

838 COVID-19 recovery can benefit biodiversity

By R. M. Pearson et al.

839 COVID-19 spotlights medical diagnostics

By M. Jankowski et al.

RESEARCH

IN BRIEF

840 From *Science* and other journals

REVIEW

843 Social environment

Social determinants of health and survival in humans and other animals *N. Snyder-Mackler et al.*

REVIEW SUMMARY; FOR FULL TEXT: [DX.DOI.ORG/10.1126/SCIENCE.AAX9553](https://doi.org/10.1126/SCIENCE.AAX9553)
VIDEO

RESEARCH ARTICLES

844 Plant science

Horizontal gene transfer of *Fhb7* from fungus underlies *Fusarium* head blight resistance in wheat *H. Wang et al.*
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: [DX.DOI.ORG/10.1126/SCIENCE.ABA5435](https://doi.org/10.1126/SCIENCE.ABA5435)

PERSPECTIVE p. 822

845 Water resources

Global threat of arsenic in groundwater *J. Podgorski and M. Berg*
PERSPECTIVE p. 818; PODCAST

850 Device technology

Aligned, high-density semiconducting carbon nanotube arrays for high-performance electronics *L. Liu et al.*

REPORTS p. 874 & 878

REPORTS

856 Topological optics

Observation of Floquet solitons in a topological bandgap *S. Mukherjee and M. C. Rechtsman*
PERSPECTIVE p. 821

860 Coronavirus

Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period *S. M. Kissler et al.*

869 Tropical forest

Long-term thermal sensitivity of Earth's tropical forests *M. J. P. Sullivan et al.*
NEWS STORY p. 813

Device technology

874 Precise pitch-scaling of carbon nanotube arrays within three-dimensional DNA nanotrenches *W. Sun et al.*

878 DNA-directed nanofabrication of high-performance carbon nanotube field-effect transistors *M. Zhao et al.*

RESEARCH ARTICLE p. 850

881 Pollinators

Bumble bees damage plant leaves and accelerate flower production when pollen is scarce *F. G. Pashalidou et al.*
PERSPECTIVE p. 824

885 Chemical physics

Simultaneous observation of nuclear and electronic dynamics by ultrafast electron diffraction *J. Yang et al.*
PERSPECTIVE p. 820

889 Metrology

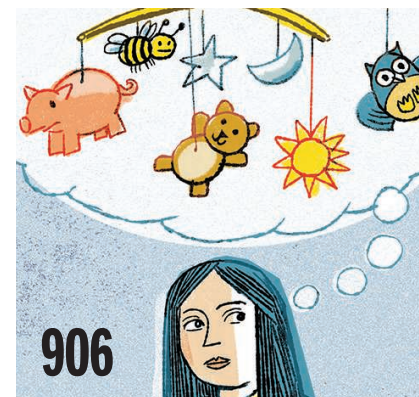
Coherent optical clock down-conversion for microwave frequencies with 10^{-18} instability *T. Nakamura et al.*
PERSPECTIVE p. 825

893 Structural biology

Ion transport and regulation in a synaptic vesicle glutamate transporter *F. Li et al.*

897 Cell biology

Supramolecular attack particles are autonomous killing entities released from cytotoxic T cells *Š. Bálint et al.*



906

DEPARTMENTS

797 Editorial

Suspend tests and rankings *By H. Holden Thorp*

906 Working Life

An end and a beginning *By Logan Brenner*

ON THE COVER



Fusarium graminearum fungus (orange) growing in a wheat spikelet. The disease *Fusarium* head blight (FHB; also known as wheat scab) threatens the world's wheat supply by reducing grain yield and quality. Researchers

have isolated a gene from wheatgrass (*Thinopyrum elongatum*), a wild relative of cultivated wheat, that decreases *Fusarium* infection and detoxifies mycotoxins in wheat grains. This gene can be introduced to wheat cultivars through breeding to minimize FHB-related losses. See pages 822 and 844. Photo: Qihui Chen/Jiayang Agricultural Bureau

Science Staff794
Science Careers902

SCIENCE (ISSN 0036-8075) is published weekly on Friday, except last week in December, by the American Association for the Advancement of Science, 1200 New York Avenue, NW, Washington, DC 20005. Periodicals mail postage (publication No. 484460) paid at Washington, DC, and additional mailing offices. Copyright © 2020 by the American Association for the Advancement of Science. The title SCIENCE is a registered trademark of the AAAS. Domestic individual membership, including subscription (12 months): \$165 (\$74 allocated to subscription). Domestic institutional subscription (61 issues): \$2148; Foreign postage extra: \$98. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request. GST #R125488122. Publications Mail Agreement Number 1069624. Printed in the U.S.A.
Change of address: Allow 4 weeks, giving old and new addresses and 8-digit account number. Postmaster: Send change of address to AAAS, P.O. Box 96178, Washington, DC 20090-6178. Single-copy sales: \$15 each plus shipping and handling available from backissues.science.org; bulk rate on request. Authorization to reproduce material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act can be obtained through the Copyright Clearance Center (CCC), www.copyright.com. The identification code for Science is 0036-8075. Science is indexed in the Reader's Guide to Periodical Literature and in several specialized indexes.

Editor-in-Chief Holden Thorp, hthorp@aaas.org

Executive Editor Monica M. Bradford

Editors, Research Valda Vinson, Jake S. Yeston Editor, Insights Lisa D. Chong

DEPUTY EDITORS Julia Fahrenkamp-Uppenbrink (UK), Stella M. Hurlley (UK), Phillip D. Szuromi, Sacha Vignieri **SR. EDITORIAL FELLOW** Andrew M. Sugden (UK) **SR. EDITORS** Gemma Alderton (UK), Caroline Ash (UK), Brent Grocholski, Pamela J. Hines, Paula A. Kiberstis, Marc S. Lavine (Canada), Steve Mao, Ian S. Osborne (UK), Beverly A. Purnell, L. Bryan Ray, H. Jesse Smith, Keith T. Smith (UK), Jelena Stajic, Peter Stern (UK), Valerie B. Thompson, Brad Wible, Laura M. Zahn **ASSOCIATE EDITORS** Michael A. Funk, Priscilla N. Kelly, Tage S. Rai, Seth Thomas Scanlon (UK), Yury V. Suleymanov **LETTERS EDITOR** Jennifer Sills **LEAD CONTENT PRODUCTION EDITORS** Harry Jach, Lauren Kmeck **CONTENT PRODUCTION EDITORS** Amelia Beyna, Jeffrey E. Cook, Chris Filatreau, Julia Katris, Nida Masiulis, Suzanne M. White **SR. EDITORIAL COORDINATORS** Carolyn Kyle, Beverly Shields **EDITORIAL COORDINATORS** Aneera Dobbins, Joi S. Granger, Jeffrey Hearn, Lisa Johnson, Maryrose Madrid, Ope Martins, Shannon McMahon, Jerry Richardson, Hilary Stewart (UK), Alana Warnke, Alice Whaley (UK), Anita Wynn **PUBLICATIONS ASSISTANTS** Jeremy Dow, Alexander Kief, Ronnel Navas, Brian White **EXECUTIVE ASSISTANT** Jessica Slater **ASI DIRECTOR, OPERATIONS** Janet Clements (UK) **ASI SR. OFFICE ADMINISTRATOR** Jessica Waldoock (UK)

News Editor Tim Appenzeller

NEWS MANAGING EDITOR John Travis **INTERNATIONAL EDITOR** Martin Enserink **DEPUTY NEWS EDITORS** Elizabeth Culotta, Lila Guterman, David Grimm, Eric Hand (Europe), David Malakoff **SR. CORRESPONDENTS** Daniel Clerly (UK), Jon Cohen, Jeffrey Mervis, Elizabeth Pennisi **ASSOCIATE EDITORS** Jeffrey Brinard, Catherine Maticic **NEWS REPORTERS** Adrian Cho, Jennifer Couzin-Frankel, Jocelyn Kaiser, Kelly Servick, Robert F. Service, Erik Stokstad, Paul Voosen, Meredith Wadman **INTERNS** Eva Frederick, Rodrigo Perez Ortega **CONTRIBUTING CORRESPONDENTS** Warren Cornwall, Ann Gibbons, Mara Hvistendahl, Sam Kean, Eli Kintisch, Kai Kupferschmidt (Berlin), Andrew Lawler, Mitch Leslie, Eliot Marshall, Virginia Morell, Dennis Normile (Shanghai), Elisabeth Pain (Careers), Charles Pillar, Michael Price, Tania Rabesandratana (Barcelona), Emily Underwood, Gretchen Vogel (Berlin), Lizzie Wade (Mexico City) **CAREERS** Donisha Adams, Rachel Bernstein (Editor), Katie Langin (Acting Editor) **COPY EDITORS** Julia Cole (Senior Copy Editor), Cyra Master (Copy Chief) **ADMINISTRATIVE SUPPORT** Meagan Welland

Creative Director Beth Rakouskas

DESIGN MANAGING EDITOR Marcy Atarod **GRAPHICS MANAGING EDITOR** Alberto Cuadra **PHOTOGRAPHY MANAGING EDITOR** William Douthitt **WEB CONTENT STRATEGY MANAGER** Kara Estelle-Powers **SENIOR DESIGNER** Chrystal Smith **DESIGNER** Christina Aycock **GRAPHICS EDITOR** Nirja Desai **INTERACTIVE GRAPHICS EDITOR** Xing Liu **SENIOR SCIENTIFIC ILLUSTRATORS** Valerie Altounian, Chris Bickel **SCIENTIFIC ILLUSTRATOR** Alice Kitterman **SENIOR GRAPHICS SPECIALISTS** Holly Bishop, Nathalie Cary **SENIOR PHOTO EDITOR** Emily Petersen **PHOTO EDITOR** Kaitlyn Dolan

Chief Executive Officer and Executive Publisher Sudip Parikh

Publisher, Science Family of Journals Bill Moran

DIRECTOR, BUSINESS SYSTEMS AND FINANCIAL ANALYSIS Randy Yi **DIRECTOR, BUSINESS OPERATIONS & ANALYSIS** Eric Knott **DIRECTOR OF ANALYTICS** Enrique Gonzales **MANAGER, BUSINESS OPERATIONS** Jessica Tierney **SENIOR BUSINESS ANALYST** Cory Lipman, Meron Kebede **FINANCIAL ANALYST** Alexander Lee **ADVERTISING SYSTEM ADMINISTRATOR** Tina Burks **SENIOR SALES COORDINATOR** Shirley Young **DIGITAL/PRINT STRATEGY MANAGER** Jason Hillman **QUALITY TECHNICAL MANAGER** Marcus Spiegler **ASSISTANT MANAGER DIGITAL/PRINT** Rebecca Doshi **SENIOR CONTENT SPECIALISTS** Steve Forrester, Jacob Hedrick, Antoinette Hodal, Lori Murphy **DIGITAL PRODUCTION MANAGER** Lisa Stanford **CONTENT SPECIALIST** Kimberley Oster **ADVERTISING PRODUCTION OPERATIONS MANAGER** Deborah Tompkins **DESIGNER, CUSTOM PUBLISHING** Jeremy Huntsinger **SR. TRAFFIC ASSOCIATE** Christine Hall **SPECIAL PROJECTS ASSOCIATE** Sarah Dhere **ASSOCIATE DIRECTOR, BUSINESS DEVELOPMENT** Justin Sawyers **GLOBAL MARKETING MANAGER** Allison Pritchard **DIGITAL MARKETING MANAGER** Aimee Aponte **JOURNALS MARKETING MANAGER** Shawana Arnold **MARKETING ASSOCIATES** Tori Velasquez, Mike Romano, Ashley Hylton **DIGITAL MARKETING SPECIALIST** Asleigh Rojanavongse **SENIOR DESIGNER** Kim Huynh

DIRECTOR AND SENIOR EDITOR, CUSTOM PUBLISHING Sean Sanders **ASSISTANT EDITOR, CUSTOM PUBLISHING** Jackie Oberst

DIRECTOR, PRODUCT & PUBLISHING DEVELOPMENT Chris Reid **DIRECTOR, BUSINESS STRATEGY AND PORTFOLIO MANAGEMENT** Sarah Whalen **ASSOCIATE DIRECTOR, PRODUCT MANAGEMENT** Kris Bishop **SR. PRODUCT ASSOCIATE** Robert Koepke **DIGITAL PRODUCT STRATEGIST** Michael Hardesty **SPJ ASSOCIATE** Samantha Bruno Fuller

DIRECTOR, INSTITUTIONAL LICENSING Iquo Edim **ASSOCIATE DIRECTOR, RESEARCH & DEVELOPMENT** Elisabeth Leonard **MARKETING MANAGER** Kess Knight **SENIOR INSTITUTIONAL LICENSING MANAGER** Ryan Rexroth **INSTITUTIONAL LICENSING MANAGER** Marco Castellani **MANAGER, AGENT RELATIONS & CUSTOMER SUCCESS** Judy Lillibridge **SENIOR OPERATIONS ANALYST** Lana Guz **FULFILLMENT COORDINATOR** Melody Stringer **SALES COORDINATOR** Josh Haverlock

DIRECTOR, GLOBAL SALES Tracy Holmes **EAST COAST AND MID WEST SALES** Stephanie O'Connor **US WEST COAST SALES** Lynne Stickrod **US SALES MANAGER, SCIENCE CAREERS** Claudia Paulsen-Young **US SALES REP, SCIENCE CAREERS** Tracy Anderson **ASSOCIATE DIRECTOR, ROW** Roger Gonçalves **SALES REP, ROW** Sarah Lelarge **SALES ADMIN ASSISTANT, ROW** Bryony Cousins **DIRECTOR OF GLOBAL COLLABORATION AND ACADEMIC PUBLISHING RELATIONS** Asia Xiaoying Chu **ASSOCIATE DIRECTOR, INTERNATIONAL COLLABORATION** Grace Yao **SALES MANAGER** Danny Zhao **MARKETING MANAGER** Kilo Lan ASCA CORPORATION, JAPAN Kaoru Sasaki (Tokyo), Miyuki Tani (Osaka) **COLLABORATION/CUSTOM PUBLICATIONS/JAPAN** Adarsh Sandhu

DIRECTOR, COPYRIGHT, LICENSING AND SPECIAL PROJECTS Emilie David **RIGHTS AND LICENSING COORDINATOR** Jessica Adams **RIGHTS AND PERMISSIONS ASSOCIATE** Elizabeth Sandler **CONTRACTS AND LICENSING ASSOCIATE** Lili Catlett

MAIN HEADQUARTERS

Science/AAAS
1200 New York Ave. NW
Washington, DC 20005

SCIENCE INTERNATIONAL

Clarendon House
Clarendon Road
Cambridge, CB2 8FH, UK

SCIENCE CHINA

Room 1004, Culture Square
No. 59 Zhongguancun St.
Haidian District, Beijing, 100872

SCIENCE JAPAN

ASCA Corporation
Sibaura TY Bldg. 4F, 1-14-5
Shibaura Minato-ku
Tokyo, 108-0073 Japan

EDITORIAL

science_editors@aaas.org

NEWS

science_news@aaas.org

INFORMATION FOR AUTHORS

sciencemag.org/authors/
science-information-authors

REPRINTS AND PERMISSIONS

sciencemag.org/help/
reprints-and-permissions

MEDIA CONTACTS

scipak@aaas.org

MULTIMEDIA CONTACTS

SciencePodcast@aaas.org
ScienceVideo@aaas.org

INSTITUTIONAL SALES

AND SITE LICENSES

sciencemag.org/librarian

PRODUCT ADVERTISING

& CUSTOM PUBLISHING
advertising.sciencemag.org/
products-services
science_advertising@aaas.org

CLASSIFIED ADVERTISING

advertising.sciencemag.org/
science-careers
advertise@sciencecareers.org

JOB POSTING CUSTOMER SERVICE

employers.sciencecareers.org
support@sciencecareers.org

MEMBERSHIP AND INDIVIDUAL

SUBSCRIPTIONS
sciencemag.org/subscriptions

MEMBER BENEFITS

aaas.org/membercentral

AAAS BOARD OF DIRECTORS

CHAIR Steven Chu
PRESIDENT Claire M. Fraser
PRESIDENT-ELECT Susan G. Amara
TREASURER Carolyn N. Ainslie
CHIEF EXECUTIVE OFFICER
Sudip Parikh
BOARD Cynthia M. Beall
Rosina M. Bierbaum
Ann Bostrom
Stephen P.A. Fodor
S. James Gates, Jr.
Laura H. Greene
Kaye Husbands Fealing
Maria M. Klawe
Robert B. Millard
Alonda Nelson
William D. Provine

BOARD OF REVIEWING EDITORS (Statistics board members indicated with \$)

Adriano Aguzzi, U. Hospital Zürich
Takuzo Aida, U. of Tokyo
Leslie Aiello, Wenner-Gren Foundation
Judith Allen, U. of Manchester
Sebastian Amigorena, Institut Curie
James Analytis, U. of California, Berkeley
Paola Ariotti, Harvard U.
Johan Auwerx, EPFL
David Awschalom, U. of Chicago
Clare Baker, U. of Cambridge
Nenad Ban, ETH Zürich
Franz Bauer, Pontificia Universidad Católica de Chile
Ray H. Baughman, U. of Texas at Dallas
Peter Bearman, Columbia U.
Carlo Beenakker, Leiden U.
Yasmine Belkaid, NIAID, NIH
Philip Benfey, Duke U.
Gabriele Bergers, VIB
Bradley Bernstein, Mass. General Hospital
Alessandra Biffi, Harvard Med. School
Peer Bork, EMBL
Chris Bowler, Ecole Normale Supérieure
Ian Boyd, U. of St. Andrews
Emily Brodsky, U. of California, Santa Cruz
Ron Brookmeyer, U. of California, Los Angeles (\$) **U. of California, Los Angeles (\$)**
Christian Büchel, UKE Hamburg
Dennis Burton, Scripps Research
Carter Tribble Butts, U. of California, Irvine
György Buzsáki, New York U. School of Med.
Blanche Capel, Duke U.
Annamarie Carlton, U. of California, Irvine
Nick Chater, U. of Warwick
Zhijian Chen, UT Southwestern Med. Ctr.
Ib Chorkendorff, Denmark TU
James J. Collins, MIT
Robert Cook-Deegan, Arizona State U.
Alan Cowman, Walter & Eliza Hall Inst.
Carolyn Coyne, U. of Pittsburgh
Roberta Croce, VU Amsterdam
Jeff L. Dangl, U. of North Carolina
Tom Daniel, U. of Washington
Chiara Daraio, Caltech
Nicolas Dauphas, U. of Chicago
Frans de Waal, Emory U.
Claude Desplan, New York U.
Sandra Diaz, Universidad Nacional de Córdoba
Ulrike Diebold, TU Wien
Hong Ding, Inst. of Physics, CAS
Jennifer Dionne, Stanford U.
Dennis Discher, U. of Penn.
Gerald Dorn, Washington U. in St. Louis
Jennifer A. Doudna, U. of California, Berkeley
Bruce Dunn, U. of California, Los Angeles
William Dunphy, Caltech
Christopher Dye, U. of Oxford
Todd Ehlers, U. of Tübingen
Jennifer Eilseisef, Johns Hopkins U.
Tim Elston, U. of North Carolina
Andrea Encalada, U. San Francisco de Quito
Nader Engheta, U. of Penn.
Karen Ersche, U. of Cambridge
Barry Everitt, U. of Cambridge
Vanessa Ezenwa, U. of Georgia
Michael Feuer, The George Washington U.
Toren Finkel, U. of Pittsburgh Med. Ctr.
Gwen Flowers, Simon Fraser U.
Peter Fratzl, Max Planck Inst. Potsdam
Elaine Fuchs, Rockefeller U.
Eileen Furlong, EMBL
Jay Gallagher, U. of Wisconsin
Daniel Geschwind, U. of California, Los Angeles
Karl-Heinz Glassmeier, TU Braunschweig
Ramon Gonzalez, U. of South Florida
Elizabeth Grove, U. of Chicago
Nicolas Gruber, ETH Zürich
Hua Guo, U. of New Mexico
Kip Guy, U. of Kentucky College of Pharmacy
Taekjip Ha, Johns Hopkins U.
Christian Haass, Ludwig Maximilians U.
Sharon Hammes-Schiffer, Yale U.
Wolf-Dietrich Hardt, ETH Zürich
Louise Harra, U. College London
Jian He, Clemson U.
Carl-Philipp Heisenberg, IST Austria
Ykä Helariutta, U. of Cambridge
Janet G. Hering, Eawag
Hans Hilgenkamp, U. of Twente
Kai-Uwe Hinrichs, U. of Bremen
Lora Hooper, UT Southwestern Med. Ctr.
Fred Hughson, Princeton U.
Randall Hulet, Rice U.
Auke Ijspeert, EPFL
Akiko Iwasaki, Yale U.
Stephen Jackson, USGS and U. of Arizona
Kai Johnson, EPFL
Peter Jonas, IST Austria
Matt Kaeblerlein, U. of Washington
William Kaelin Jr., Dana-Farber Cancer Inst.
Daniel Kammen, U. of California, Berkeley
V. Naray Kim, Seoul Nat. U.
Robert Kingston, Harvard Med. School
Nancy Knowlton, Smithsonian Institution
Etienne Koechlin, Ecole Normale Supérieure
Alexander L. Kolodkin, Johns Hopkins U.
Julija Krupic, U. of Cambridge
Thomas Langer, Max Planck Inst. Cologne
Mitchell A. Lazar, U. of Penn.
Ottoline Leyser, U. of Cambridge
Wendell Lim, U. of California, San Francisco
Jiangui Liu, Michigan State U.
Luis Liz-Marzán, CIC biomaGUNE
Jonathan Losos, Washington U. in St. Louis
Ke Lu, Chinese Acad. of Sciences
Christian Lüscher, U. of Geneva
Jean Lynch-Stieglitz, Georgia Inst. of Tech.
Fabienne Mackay, U. of Melbourne
Anne Magurran, U. of St. Andrews
Oscar Marín, King's College London
Charles Marshall, U. of California, Berkeley
Christopher Marx, U. of Idaho
Geraldine Masson, CNRS
C. Robertson McClung, Dartmouth College
Rodrigo Medellín, U. Nacional Autónoma de México
Graham Medley, London School of Hygiene & Tropical Med.
Jane Memmott, U. of Bristol
Baoxia Mi, U. of California, Berkeley
Edward Miguel, U. of California, Berkeley
Tom Misteli, NCI, NIH
Yasushi Miyashita, U. of Tokyo
Alison Motingser-Reif, NIEHS, NIH (\$) **NIH (\$)**
Daniel Nettle, Newcastle U.
Daniel Neumark, U. of California, Berkeley
Beatriz Noheida, U. of Groningen
Helga Nowotny, Vienna Science, Research & Tech. Fund
Rachel O'Reilly, U. of Birmingham
Harry Orr, U. of Minnesota
Pilar Ossorio, U. of Wisconsin
Andrew Oswald, U. of Warwick
Isabella Pagano, Istituto Nazionale di Astrofisica
Margaret Palmer, U. of Maryland
Elizabeth Levy Paluck, Princeton U.
Jane Parker, Max Planck Inst. Cologne
Giovanni Parmigiani, Dana-Farber Cancer Inst. (\$) **Dana-Farber Cancer Inst. (\$)**
Samuel Pfaff, Salk Inst. for Biological Studies
Julie Pfeiffer, UT Southwestern Med. Ctr.
Matthieu Piel, Institut Curie
Kathrin Plath, U. of California, Los Angeles
Martin Pollard, U. of Penn.
Katherine Pollard, U. of California, San Francisco
Elvira Poloczanska, Alfred-Wegener-Inst.
Julia Pongratz, Ludwig Maximilians U.
Philippe Poulin, CNRS
Jonathan Pritchard, Stanford U.
Félix A. Rey, Institut Pasteur
Trevor Robbins, U. of Cambridge
Joeri Rogelj, Imperial College London
Amy Rosenzweig, Northwestern U.
Mike Ryan, U. of Texas at Austin
Mitinori Saitou, Kyoto U.
Shimon Sakaguchi, Osaka U.
Miquel Salmeron, Lawrence Berkeley Nat. Lab
Nitin Samarth, Penn. State U.
Jürgen Sandkühner, Med. U. of Vienna
Alexander Schier, Harvard U.
Wolfram Schlenker, Columbia U.
Susannah Scott, U. of California, Santa Barbara
Rebecca Sear, London School of Hygiene & Tropical Med.
Vladimir Shalae, Purdue U.
Jie Shan, Cornell U.
Beth Shapiro, U. of California, Santa Cruz
Jay Shendure, U. of Washington
Steve Sherwood, U. of New South Wales
Brian Shoichet, U. of California, San Francisco
Robert Siliciano, Johns Hopkins U. School of Med.
Lucia Sivilotti, U. College London
Alison Smith, John Innes Centre
Richard Smith, U. of North Carolina (\$) **U. of North Carolina (\$)**
Mark Smyth, QIMR Berghofer
Pam Soltis, U. of Florida
John Speakman, U. of Aberdeen
Tara Spire-Jones, U. of Edinburgh
Allan C. Spradling, Carnegie Institution for Science
V. S. Subrahmanian, Dartmouth College
Ira Tabas, Columbia U.
Sarah Teichmann, Wellcome Sanger Inst.
Rocio Titiunik, Princeton U.
Shubha Tole, Tata Inst. of Fundamental Research
Wim van der Putten, Netherlands Inst. of Ecology
Reinhold Veugelers, KU Leuven
Bert Vogelstein, Johns Hopkins U.
Kathleen Vohs, U. of Minnesota
David Wallace, Weizmann Inst. of Science
Jane-Ling Wang, U. of California, Davis (\$) **U. of California, Davis (\$)**
David Waxman, Fudan U.
Jonathan Weissman, U. of California, San Francisco
Chris Winkle, U. of Missouri (\$) **U. of Missouri (\$)**
Terrie Williams, U. of California, Santa Cruz
Ian A. Wilson, Scripps Research (\$) **Scripps Research (\$)**
Yu Xie, Princeton U.
Jan Zaanen, Leiden U.
Kenneth Zaret, U. of Penn. School of Med.
Jonathan Zehr, U. of California, Santa Cruz
Xiaowei Zhuang, Harvard U.
Maria Zuber, MIT

Suspend tests and rankings

The notion that U.S. colleges and universities will open this fall in “normal” mode should not be in any forecast. As Dr. Anthony Fauci (of the White House’s coronavirus task force) testified last week before the Senate, it’s unlikely that a vaccine or treatments for coronavirus disease 2019 (COVID-19) will be available by the time students return to campus. I sympathize with the predicament of college and university administrators who need to reopen in ways that are safe and supportive for all their students while also planning for the possibility that they won’t be able to reopen in-person classes. To help them grapple with this, let’s suspend two things, at least temporarily—test scores and rankings.

Although universities that support research, graduate, and postgraduate training have struggled during the pandemic with the shuttering of labs, clinics, and academic programs, these functions seem to be on their way back and probably can restart safely. I worry less about the recovery of this sector of higher education than I do about undergraduate students, of which there are an estimated 20 million in the United States. We know that their success, on multiple fronts, is enhanced by completing college.

Recent statements by a few U.S. college and university presidents about the coming fall semester range from the bullish announcement that Purdue University will open with in-person classes to the cautious decision that the California State University system will be all-virtual. My guess is that there will be a messy, hybrid solution involving mainly virtual instruction, for most institutions. In-person classes will require new configurations for housing and dining, smaller lectures, and more instructors. Student health centers need to prepare for testing, isolation, and mental health support. And those are just the most obvious needs to be addressed, quickly. The chaotic move to virtual classes this spring demonstrated that this approach needs to be executed much more deliberately in the fall, which will require resources to help faculty prepare for a new mode of teaching.

My biggest worry is that certain students may get lost in the planning debates and that COVID-19 health and economic impacts may further exacerbate inequities in higher education. The spring semester showed us that students had to make quick arrangements to continue

their education online—a path that was easier for some than others. And the large number of students who already lived off campus—particularly those enrolled in community colleges and big urban public universities—were in the same situation as they were in the pre-pandemic era, but without adequate recognition. Shutting down in-person classes and campuses all together compounded student insecurities—from food, shelter, and medical to financial and technological. A major concern is whether these students will be able to continue (or even begin) their higher education in the fall.

For institutional leaders strategizing to reopen, addressing the imbalances in college access, enrollment, and completion of undergraduate education should be a priority. High scores in admissions tests and high ability to pay tuition are already given too much weight by American academic institutions when it comes to

undergraduate admissions. This inequitable behavior is further reinforced by the yearly rankings assigned to colleges and universities, most notoriously by *U.S. News and World Report* (since 1983), which university donors and political stakeholders study more than they should. To any logical scientific observer, the fine distinctions of where schools show up on this list are statistically meaningless—but try telling that to a roomful of

alumni or parents. Countless hours of trustee meetings are spent going over the minute details of the formula and setting institutional goals. Achieving these goals usually means doing things that make the college or university less accessible, like admitting more students with high standardized test scores.

A truly transformative move in this moment of crisis would be to suspend testing requirements and college rankings. This is not a time for undergraduate institutions to be using precious resources to chase these numbers. Rather, they need to support struggling students and other members of the academic community so that education can resume this fall in a manner that is fair to all. Some schools are already making test scores optional for the time being, and hopefully that requirement will never return. Ranking colleges and universities changed higher education, mostly for the worse. Now is the time for institutions to suspend those rankings and, when the crisis is over, bring them back in a more progressive form.

—H. Holden Thorp



H. Holden Thorp
Editor-in-Chief,
Science journals.
hthorp@aaas.org;
@hholdenthorp

“...COVID-19...may
further exacerbate
inequities in
higher education.”

NEWS



IN BRIEF

Edited by Jeffrey Brainard

DISPATCHES FROM THE PANDEMIC

Patrons enjoy drinks inside a bar in Bloomington, Indiana, on 16 May after authorities there and in many other U.S. jurisdictions lifted stay-at-home orders.

Warp Speed leaders are named

VACCINES | The White House on 15 May named a veteran of vaccine development and a U.S. Army general to lead a crash, \$10 billion program to develop vaccines, drugs, and diagnostics for COVID-19. It aims to deliver one or more vaccines to 300 million Americans by January 2021, a nationalistic goal potentially at odds with a call last week by more than 140 world leaders to equitably distribute any vaccine globally. Moncef Slaoui, who formerly led global vaccine development at GlaxoSmithKline, will guide the science for the U.S. effort, called Operation Warp Speed, while Gen. Gustave Perna will oversee the supply chain and delivery of products. The private-public partnership plans to vaccinate Americans first before sharing products with other countries. The U.K. government took a similar stand last week about a vaccine being developed with its funding by the University of Oxford. And French vaccine manufacturer Sanofi, under pressure from French leaders, retreated from a claim by its CEO last week that it would supply the United States first because of early funding from the U.S. government.

Trump eyes pulling out of WHO

POLITICS | Public health researchers criticized President Donald Trump's threat this week to end U.S. membership in the World Health Organization (WHO) and permanently withdraw funding. Trump has said China did not act quickly enough to stop the COVID-19 pandemic; in an 18 May letter to WHO, he complained that it, too, moved too slowly and failed to confront China. Scientists said his statements were unjustified and contained factual inaccuracies about WHO's role and how the pandemic unfolded. "China and the U.S. are fighting it out like divorced parents while WHO is the child caught in the middle trying not to pick sides," said Devi Sridhar, a professor of global public health at the University of Edinburgh.

CDC reopening guidance faulted

POLICY | The U.S. Centers for Disease Control and Prevention (CDC) drew controversy last week by releasing a set of recommendations, which critics complained were watered down, to guide the country's reopening from COVID-19 lockdowns. The guidelines replaced an earlier draft that White House officials

had rejected as too prescriptive and burdensome on businesses. The replacement consists of six flow charts describing safeguards such as cleaning measures and employee monitoring that schools, offices, and other facilities should begin before reopening. Public health experts, including Anthony Fauci, head of the U.S. National Institute of Allergy and Infectious Diseases and a member of the White House Coronavirus Task Force, continued to warn of disastrous consequences of easing lockdowns prematurely. By this week, all but a handful of states had reopened retail stores, and many began to allow dine-in service to resume at restaurants, though nearly all schools remain closed for the academic year.

China's halt cut pollution deaths

PUBLIC HEALTH | Even as COVID-19 took thousands of lives in China, a reduction in air pollution from the country's lockdown saved more lives than those lost to the novel coronavirus, a study has estimated. The quarantine reduced traffic, leading to a 37% decline in emissions of nitrogen dioxide and a 30% fall in fine particulate matter at the height of China's quarantine, from 10 February to 14 March. This prevented

an estimated 12,151 deaths from diseases, such as stroke, linked to these pollutants, compared with 4633 deaths attributed to the virus, Yale University researchers reported online on 13 May in *The Lancet*. The findings suggest substantial health benefits can arise from aggressive control measures for air pollution, the authors wrote. Other researchers have reported that air pollution fell during COVID-19 lockdowns. A preprint study by a Norwegian-German group, for example, reported reductions in premature deaths and pediatric asthma in China and India because of reduced traffic.

Few Spaniards have antibodies

EPIDEMIOLOGY | A study of blood samples from 70,000 people in Spain suggests only a small minority have antibodies to the virus that causes COVID-19, despite the country having had the fifth most deaths from the disease. The finding implies the population is nowhere near a level of immunity that would slow its spread. At a 13 May news conference, Spanish health officials announced preliminary results from their nationwide study of 36,000 randomly selected households, which aimed to find out how many people have been infected and already recovered, one of the largest such studies in the world to date. They estimated that roughly 2.3 million Spaniards, about 5% of the population, have antibodies that might protect against a future infection. The numbers varied widely in different regions: In hard-hit Madrid, just over 11% of people tested positive. In Barcelona it was about 7% but less than 2% in several southern regions.

More child deaths forecast

GLOBAL HEALTH | The COVID-19 pandemic is disrupting routine health services and food supplies in ways that could kill an additional 1.2 million children under age 5 and 57,000 mothers worldwide over the next 6 months, a study estimates. The toll for children would represent a 45% increase over the level expected during normal times. Mortality rates for COVID-19 itself appear to be low in children and in women of reproductive age, but the disruption caused by the pandemic is hampering the delivery of vaccines as well as treatments for infectious diseases and obstetrics care, the authors reported 12 May online in *The Lancet Global Health*. That could reverse a 50% decline since 2000 in death rates among children under 5.

SCIENCEMAG.ORG/TAGS/CORONAVIRUS

Read additional *Science* coverage of the pandemic.



A plant specialist at Monsanto (now Bayer) hand pollinates genetically modified corn plants.

AGRICULTURE

U.S. eases rules for bioengineered plants

A major change to U.S. regulation of biotech will exempt some gene-edited plants from government oversight. The new policy, published in the *Federal Register* on 18 May, says that if researchers use gene editing to design a plant that could have been bred conventionally, the new plant will be exempt from regulation. But anything else—such as moving a gene between species—will still require a regulatory review. In addition, the government will cease regulating new varieties of already approved GM crops, easing their path to market. Industry groups are welcoming the new rule, whereas opponents are decrying the reduction of government oversight. Most of the changes go into effect 5 April 2021.

NIH head lauded for faith ties

PRIZES | U.S. National Institutes of Health (NIH) Director Francis Collins has won the \$1.3 million 2020 Templeton Prize for his work to reconcile science and religion. Collins, 70, who embraced Christianity as a medical student, led the Human Genome Project to its completion in 2003 before taking the helm of NIH 11 years ago. His 2006 book, *The Language of God: A Scientist Presents Evidence for Belief*, became a bestseller, and he continues to speak publicly about his faith, despite criticism from some scientists who find such activities inappropriate for the leader of a federal science agency. Collins “has demonstrated how religious faith can motivate and inspire rigorous scientific research,” the John Templeton Foundation said in a release. The prize’s judges selected Collins last year, before the COVID-19

pandemic; since it began, he “has urged faith communities to trust science” to fight the disease, the organization added. The John Templeton Foundation funds research as well as projects at the intersection of science and religion (including at AAAS, which publishes *Science*).

A boost for geochronology

EARTH SCIENCE | A blue-ribbon panel this week recommended creating a new consortium for geochronology, which dates the age of rocks, to improve coordination and funding. The idea came in a set of recommendations released this week by the National Academies of Sciences, Engineering, and Medicine (NASEM) for the National Science Foundation’s (NSF’s) \$180 million annual budget for earth science this decade. At a cost of up to \$10 million annually, the geochronology consortium would

develop new dating techniques while providing dates as a service to NSF-funded researchers. The NASEM report also recommends building a \$3 million anvil press for simulating the pressures in Earth's core and creating a center for geophysicists who study hydrology and other phenomena close to the planet's surface. It endorsed development of two initiatives that would require new funding: one to study subduction zones, where oceanic crust dives beneath continents, and the other to map the continent's "critical zone," the subsurface layer of soil, rocks, and water that fuels life above.

New head of U.K. funding agency

LEADERSHIP | Ottoline Leyser, a plant biologist at the University of Cambridge,

will be the next director of the United Kingdom's £7 billion research funding agency, UK Research and Innovation (UKRI), the government announced last week. Leyser directs the Sainsbury Laboratory at Cambridge, a leading plant research center with about 130 scientists. She serves on the prime minister's Council for Science and Technology and has called for increasing diversity in science and improving research culture. UKRI will see large budget increases as the government tries to raise R&D investment to 2.4% of gross domestic product over 7 years. But the implementation of Brexit in December has left unanswered questions about U.K. participation in the European Union's research funding programs. Leyser takes over from Mark Walport next month.

EPA skips regulating pollutant

ENVIRONMENT | In the latest twist in a decades-old dispute, the U.S. Environmental Protection Agency has reportedly decided it will not set a drinking water standard for perchlorate, a pollutant linked to brain damage. The chemical, which is used in rocket fuel and other products but also occurs naturally, has seeped into numerous water supplies. Environmentalists have long pushed for its federal regulation, and in 2011 the Obama administration began that process. But the Trump administration has decided the chemical does not pose a serious threat and does not require federal oversight because some states already regulate it, *The New York Times* reported on 14 May.

THREE Qs



Witness to a landscape's recovery

Charlie Crisafulli first visited Mount St. Helens 2 months after the 18 May 1980 eruption that ripped the top off the volcano, killed 57 people, and destroyed 600 square kilometers of forests. He was a 22-year-old with an undergraduate degree in ecology, accompanying other scientists who swarmed to the mountain. Since then, Crisafulli—now an ecologist with the U.S. Forest Service—has spent much of each summer taking the mountain's pulse as life returns. (Read a longer version of this interview at <https://scim.ag/CrisafulliQA>.)

Q: What did you see after the eruption?

A: We were all absolutely blown away by the intensity and extent of the disturbance. Our design was to take advantage of the natural experiment that the volcano created because there were different intensities of disturbance. ... I saw this one sprig, and it was a parsley fern. Every spring since then I have gone and said hello, greeted that same exact plant.

Q: What are some of the key scientific insights from the work?

A: The initial impression was that the regeneration of the area's ecology is going to come from the edges and from distant source populations. Instead what we found is that, in some 90% of the landscape, the rule was survivorship, albeit at greatly reduced numbers and in isolated refugias. And that was not something we anticipated.

Q: How has the research informed understanding of other volcanoes?

A: When a volcano starts acting up in Chile, or in Japan, or Iceland, or New Zealand, we're often contacted to say, "Help us anticipate what's likely to happen." What we have seen is strikingly similar patterns of ecological response. ... In all cases, what we found is nutrient-impooverished volcanic material. And that's one of the biggest obstacles [to ecological recovery]. There seems to be always heroes at these volcanoes that play critically important roles. At Mount St. Helens, it was species such as lupine and alder. They had a special association on their roots with bacteria that produce nitrogen. So they were able to exploit these landscapes, and in doing so, they modified these sites and facilitated the colonization by many other plants and animals.



Cabotegravir provided long-term protection from HIV when injected intramuscularly.

IN DEPTH

BIOMEDICINE

Long-acting drug acts like a short-term AIDS vaccine

Simpler than daily pills, injections to prevent HIV infections protect people for 2 months

By **Jon Cohen**

It's not an AIDS vaccine, but it may be the closest thing to one so far. A long-acting antiretroviral drug given as an injection every 2 months powerfully protected uninfected people from HIV in a large-scale clinical trial. Although it was cut short by the COVID-19 pandemic and has not been published in a peer-reviewed journal or presented at a meeting, the study holds out hope of preventing HIV infections in high-risk groups without the need to take pills every day, which many people find difficult.

The trial compared the experimental drug cabotegravir with Truvada, a daily pill combining two antiretroviral compounds that is the standard regimen for pre-exposure prophylaxis (PrEP), as the strategy of preventing infection with drugs is known. Sponsored by the U.S. National Institute of Allergy and Infectious Diseases (NIAID), the trial began in December 2016 and enrolled more than 4500 men who have sex with men and transgender women. Participants were randomly assigned to Truvada, intramuscular injections of cabotegravir, placebo pills, or dummy injections. As of late April, 12 infections occurred in the cabotegravir group versus

38 in the equally sized group that received Truvada: a 0.38% incidence in the cabotegravir arm versus 1.21% in the Truvada one. This overall result, reported by press releases on 18 May, indicates cabotegravir works just as well as Truvada, and perhaps better, though the trial saw too few infections for that conclusion to reach statistical significance.

"It's really exciting," says Jared Baeten, an epidemiologist at the University of Washington, Seattle, who conducted a landmark PrEP study of anti-HIV pills—including Truvada—in Kenya and Uganda. "It gives another option for people who can't or don't want to take daily pills." That may get more people to try PrEP, Baeten adds, noting that more women began to use contraception as options increased. Epidemiologist Kevin De Cock, who heads the Kenya program for the U.S. Centers for Disease Control and Prevention, says, "A new vista of HIV prevention has been opened."

Several PrEP studies with Truvada have shown that, by and large, people who became infected with HIV did not take the daily pills. (Blood measurements can reveal levels of the drugs.) To address that problem, "People are looking aggressively for long-acting agents," says Myron Cohen, a microbiologist

at the University of North Carolina School of Medicine, who is co-principal investigator of the NIAID HIV Prevention Trials Network. Cabotegravir, which has worked well as a treatment in clinical trials, is further along in development than other long-acting drugs. Made by ViiV Healthcare, a joint venture of GlaxoSmithKline, Pfizer, and Shionogi that focuses on AIDS drugs, the compound targets integrase, an HIV protein that is essential to its replication.

COVID-19 shut down 11 of the trial's 43 sites in seven countries, and participants at others increasingly had trouble attending appointments, says Raphael Landovitz of the University of California, Los Angeles, the study's protocol chair. The investigators told the study's independent data and safety monitoring board (DSMB) about this problem in early April and asked that the end-point of the trial be changed to make sure the study had an "undisrupted data set." Instead of evaluating whether the injections were superior to the Truvada pills—the original goal—they decided to assess a lower standard of "noninferiority."

On 14 May, the DSMB members took a scheduled interim peek at the data. They found that the cabotegravir arm met that threshold and that the injections appeared safe and mostly well-tolerated. Study organizers now plan to offer the injections to all participants. "With cabotegravir lasting at least a couple of months we're moving in the direction of providing people with some of the benefits of a vaccine," Cohen says.

Landovitz says ongoing analyses should reveal why people who received shots of the drug became infected. Earlier studies found that men and transgender women who are underweight more rapidly eliminate the drug from their blood. Another possibility is that participants who got infected may have missed some doses of the cabotegravir pills they had to take for 5 weeks—for safety reasons—before starting the injections. Or someone could have been infected with an HIV variant that is resistant to cabotegravir.

Injected cabotegravir is also being tested in a large study of heterosexual women, which could have clear data as early as November. ViiV said in a statement it is "actively working with global health partners" to make sure that people in resource-limited countries have access to cabotegravir injections, but it did not reveal any information about pricing. Cohen says the addition of long-acting cabotegravir as PrEP provides a powerful new HIV prevention weapon. "This agent really could contribute to the goal of ending AIDS by 2030." ■



Large numbers of people working close together in a cold environment may make meatpacking plants fertile ground for the novel coronavirus.

COVID-19

Case clustering emerges as key pandemic puzzle

Why do some patients infect many others, whereas many don't spread the virus at all?

By Kai Kupferschmidt

When 61 people met for a choir practice in a church in Mount Vernon, Washington, on 10 March, everything seemed normal. For 2.5 hours the chorists sang, snacked on cookies and oranges, and sang some more. But one of them had been suffering for 3 days from what felt like a cold—and turned out to be COVID-19. In the following weeks, 53 choir members got sick, three were hospitalized, and two died, according to a 12 May report by the U.S. Centers for Disease Control and Prevention (CDC) that meticulously reconstructed the tragedy.

Many similar “superspreading events” have occurred in the COVID-19 pandemic. A database by Gwenan Knight and colleagues at the London School of Hygiene & Tropical Medicine (LSHTM) lists an outbreak in a dormitory for migrant workers in Singapore linked to almost 800 cases; 80 infections tied to live music venues in Osaka, Japan; and a cluster of 65 cases resulting from Zumba classes in South Korea. Clusters have also occurred aboard ships and at nursing homes, meatpacking plants, ski resorts, churches, restaurants, hospitals, and

prisons. Sometimes a single person infects dozens of people, whereas other clusters unfold across several generations of spread, in multiple venues.

Other infectious diseases also spread in clusters. But COVID-19, like two of its cousins, severe acute respiratory syndrome (SARS) and Middle East respiratory syndrome (MERS), seems especially prone to attacking groups of tightly connected people while sparing others. It's an encouraging finding, scientists say, because it suggests that restricting gatherings where super-spreading is likely to occur will have a major impact on transmission and that other restrictions—on outdoor activity, for example—might be eased.

“If you can predict what circumstances are giving rise to these events, the math shows you can really, very quickly curtail the ability of the disease to spread,” says Jamie Lloyd-Smith of the University of California, Los Angeles, who has studied the spread of many pathogens. But super-spreading events are ill-understood and difficult to study, and the findings can lead to heartbreak and fear of stigma in patients who touch them off.

Most of the discussion around the spread of SARS-CoV-2 has concentrated on the average number of new infections caused by each patient. Without social distancing, this reproduction number (R) is about three. But in real life, some people infect many others and others don't spread the disease at all. In

fact, the latter is the norm, Lloyd-Smith says: “The consistent pattern is that the most common number is zero. Most people do not transmit.”

That's why in addition to R , scientists use a value called the dispersion factor (k), which describes how much a disease clusters. The lower k is, the more transmission comes from a small number of people. In a seminal 2005 *Nature* paper, Lloyd-Smith and co-authors estimated that SARS—in which superspreading played a major role—had a k of 0.16. The estimated k for MERS, which emerged in 2012, is about 0.25. In the flu pandemic of 1918, in contrast, the value was about one, indicating that clusters played less of a role.

Estimates of k for SARS-CoV-2 vary. In January, researchers at the University of Bern simulated the epidemic in China for different combinations of R and k and com-

Science's
COVID-19
coverage
is supported
by the
Pulitzer Center.

pared the outcomes with what had actually taken place. They concluded that *k* for COVID-19 is somewhat higher than for SARS and MERS. But in a March preprint, Adam Kucharski of LSHTM estimated it's only 0.1. "Probably about 10% of cases lead to 80% of the spread," Kucharski says.

If he is right, SARS-CoV-2 needs to be introduced undetected into a new country at least four times to have an even chance of establishing itself, Kucharski says. That may explain why the virus did not take off around the world sooner after it emerged in China, and why some very early cases elsewhere—such as one in France in late December 2019, reported on 3 May—apparently failed to ignite a wider outbreak. If the Chinese epidemic was a big fire that sent sparks flying around the world, most of the sparks simply fizzled out.

Why coronaviruses cluster so much more than other pathogens is "a really interesting open scientific question," says Christophe Fraser of the University of Oxford, who has studied superspreading in Ebola and HIV. Their mode of transmission may be one factor. SARS-CoV-2 appears to transmit mostly through droplets, but it does occasionally spread through finer aerosols that can stay suspended in the air, enabling one person to infect many. Most published large transmission clusters "seem to implicate aerosol transmission," Fraser says.

Individual patients' characteristics play a role as well. Some people shed far more virus, and for a longer period of time, than others, perhaps because of differences in their immune system or the distribution of virus receptors in their body. A 2019 study of healthy people showed some breathe out many more particles than others when they talk. (The volume at which they spoke explained some of the variation.) Singing may release more virus than speaking, which could help explain the choir outbreaks. People's behavior also plays a role. Having many social contacts or not washing your hands makes you more likely to pass on the virus.

Superspreading usually happens indoors. Researchers in China studying the spread of the coronavirus outside Hubei province—ground zero for the pandemic—identified 318 clusters of three or more cases between 4 January and 11 February, only one of which originated outdoors. A study in Japan found that the risk of infection indoors is almost 19 times higher than outdoors. Some situations may be particularly risky. Meatpacking plants are likely vulnerable because many people work closely together in spaces where

low temperature helps the virus survive.

Countries that have beaten back the virus to low levels need to be especially vigilant for superspreading events, because they can easily undo hard-won gains. After South Korea relaxed social distancing rules in early May, a man who later tested positive for COVID-19 visited several clubs in Seoul; public health officials scrambled to identify thousands of potential contacts and have already found 170 new cases.

If public health workers knew where clusters are likely to happen, they could try to prevent them and avoid shutting down broad swaths of society, Kucharski says. "Shutdowns are an incredibly blunt tool," he says. "You're basically saying: We don't know enough about where transmission is happening to be able to target it, so we're just going to target all of it."

But studying large COVID-19 clusters is harder than it seems. Many countries have not collected the kind of detailed contact tracing data needed. And the shutdowns have been so effective that they also robbed researchers of a chance to study superspreading events. (Before the shutdowns, "there was probably a 2-week window of opportunity when a lot of these data could have been collected," Fraser says.)

The research is also prone to bias. People are more likely to remember attending a basketball game than, say, getting a haircut, a phenomenon called recall bias that may make clusters seem bigger than they are. Clusters that have an interesting social angle—such as prison outbreaks—may get more media coverage and thus jump out to researchers, while others remain hidden. Clusters of mostly asymptomatic infections may be missed altogether.

Privacy is another concern. Untangling the links between patients can reveal who was at the origin of a cluster or expose information about people's private lives. In its report about the chorus, CDC left out a seating map that could show who brought the virus to the practice. Some clubs involved in the new South Korean cluster were gay venues, which resulted in an antigay backlash and made contact tracing harder.

Fraser, who is tracking HIV transmission in Africa by sequencing virus isolates, says it is a difficult trade-off, but one that can be managed through good oversight and engagement with communities. Epidemiologists have "a duty" to study clusters, he says: "Understanding these processes is going to improve infection control, and that's going to improve all of our lives." ■

"Probably about 10% of cases lead to 80% of the spread."

Adam Kucharski,
London School of Hygiene
& Tropical Medicine

COVID-19

T cells found in coronavirus patients 'bode well' for long-term immunity

New findings suggest past infections may offer some protection against the novel coronavirus

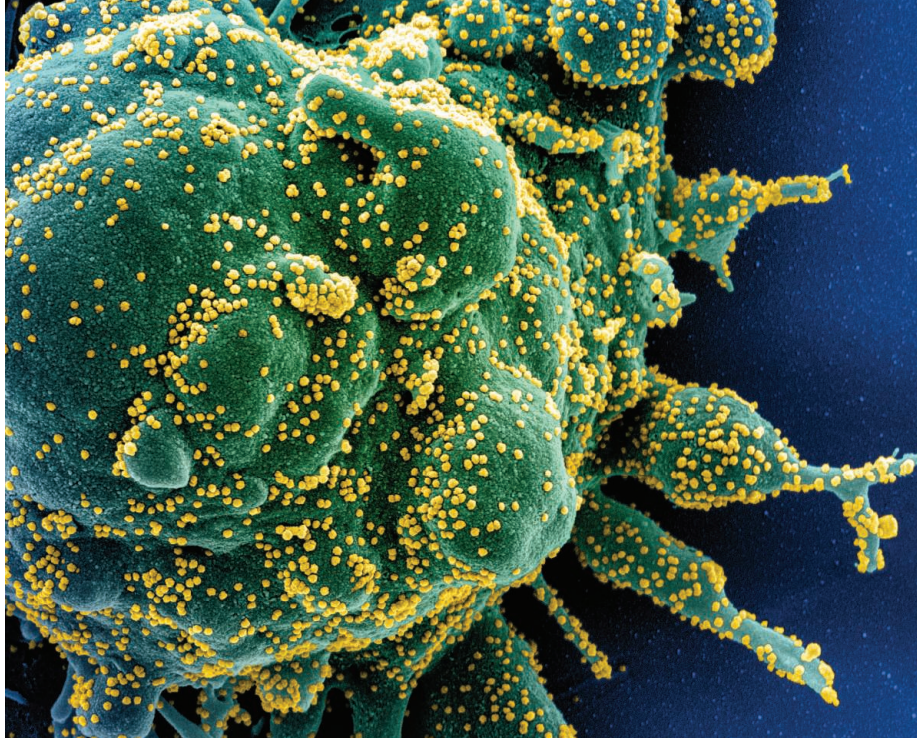
By **Mitch Leslie**

T cells are among the immune system's most powerful weapons, but their importance for battling SARS-CoV-2, the virus that causes COVID-19, has been unclear. Now, two studies show infected people harbor T cells that target the virus—and may help them recover. Both studies also found that some people never infected with SARS-CoV-2 have these cellular defenses, most likely because they were previously infected with other coronaviruses that cause the common cold.

"This is encouraging data," says virologist Angela Rasmussen of Columbia University, who wasn't involved in the work. Although the studies don't clarify whether people who clear a SARS-CoV-2 infection can ward off the virus in the future, both identified strong T cell responses to it, which "bodes well for the development of long-term protective immunity," Rasmussen says. The findings could also help researchers create better vaccines.

The more than 100 COVID-19 vaccines in development mainly focus on triggering a different immune response: antibodies. Researchers know our B cells make antibodies against SARS-CoV-2, which vaccine developers hope can latch onto the virus and prevent it from entering cells. But T cells can also help thwart infections. Helper T cells spur B cells and other immune defenders into action, whereas killer T cells target and destroy infected cells. The severity of disease can depend on the strength of these T cell responses.

To determine whether the new coronavirus provokes T cells, a team led by Shane



Immune hunters called T cells can seek and destroy a cell (green) infected with SARS-CoV-2 (yellow).

Crotty and Alessandro Sette, immunologists at the La Jolla Institute for Immunology, used bioinformatics tools to predict which segments of the virus' proteins should stimulate T cells most effectively. They then exposed immune cells from 10 patients who had recovered from mild cases of COVID-19 to these viral snippets.

All of the patients carried helper T cells that recognized the SARS-CoV-2 spike protein, which enables the virus to infiltrate our cells. They also harbored helper T cells that reacted to other SARS-CoV-2 proteins. And in 70% of the subjects, the team detected virus-specific killer T cells, they reported last week in *Cell*. "The immune system sees this virus and mounts an effective immune response," Sette says.

The results jibe with those of a study posted as a preprint on medRxiv on 22 April by immunologist Andreas Thiel of the Charité University Hospital in Berlin and colleagues. They identified helper T cells targeting the spike protein in 15 out of 18 patients hospitalized with COVID-19.

Before these studies, researchers didn't know whether T cells played a role in eliminating SARS-CoV-2—or even whether they could provoke a dangerous immune system overreaction. "These papers are really helpful because they start to define the T cell component of the immune response," Rasmussen says.

She and other scientists caution that the results do not mean people who have recovered from COVID-19 are safe from reinfection. But they do raise hopes that a vaccine could give lasting protection against the virus. To spark production of antibodies, a vaccine needs to stimulate helper T cells,

Crotty notes. "It is encouraging that we are seeing good helper T cell responses against SARS-CoV-2 in COVID-19 cases," he says.

The results have other significant implications for vaccine design, says molecular virologist Rachel Graham of the University of North Carolina, Chapel Hill. Most vaccines under development aim to elicit an immune response against the spike protein, but the La Jolla group's study determined that T cells reacted to several viral proteins, suggesting vaccines that incite an immune response to these proteins as well could be more effective. "It is important to not just concentrate on one protein," Graham says.

Both teams also wondered whether people who haven't been infected with SARS-CoV-2 also produce T cells that could combat it. Thiel and colleagues analyzed blood from 68 uninfected people and found that 34% hosted helper T cells that recognized SARS-CoV-2. The La Jolla team studied stored blood samples collected between 2015 and 2018, well before the current pandemic began, and detected these cross-reactive helper T cells in about half of them. The researchers think these cells were likely triggered by past infection with one of the four human coronaviruses that cause colds; proteins in these viruses resemble those of SARS-CoV-2.

The studies don't show people with this cross-reactivity are less likely to become ill from COVID-19. But viral immunologist Steven Varga of the University of Iowa says the results do suggest "one reason that a large chunk of the population may be able to deal with the virus is that we may have some small residual immunity from our exposure to common cold viruses." ■

COVID-19

AI systems aim to sniff out coronavirus outbreaks

Alerts would not substitute for on-the-ground sleuthing, experts warn

By **Adrian Cho**

The international alarm about the COVID-19 pandemic was sounded first not by a human, but by a computer. HealthMap, a website run by Boston Children's Hospital, uses artificial intelligence (AI) to scan social media, news reports, internet search queries, and other data for signs of disease outbreaks. On 30 December 2019, it spotted a news report of a new type of pneumonia in Wuhan, China, and issued a one-line email bulletin that seven people were in critical condition, rating the urgency at three on a scale of five.

Humans weren't far behind. Colleagues in Taiwan had already alerted Marjorie Pollack, a medical epidemiologist in New York City, to social media chatter in China that reminded her of the 2003 outbreak of severe acute respiratory syndrome (SARS), which spread to dozens of countries and killed 774. "It fit all of the been there, done that déjà vu for SARS," Pollack says. Less than 1 hour after the HealthMap alert, she posted a more detailed notice to ProMED, a list server with 85,000 subscribers for which she is a deputy editor.

But the early alarm from HealthMap underscores the potential of AI, or machine learning, to keep watch for contagion. As the COVID-19 pandemic continues to spread, AI researchers are teaming with tech companies to build automated tracking systems that will mine social media, news reports, and public health data for signs of new outbreaks. AI is no substitute for traditional public health monitoring, cautions Matthew Biggerstaff, an epidemiologist with the U.S. Centers for Disease Control and Prevention (CDC). "This should be viewed as one tool in the toolbox," he says. But it can fill a need, says Elad Yom-Tov, a Microsoft computer scientist who has worked with public health officials in the

United Kingdom. “There’s such a wealth of data, we will need some sort of tool ... to me that tool is machine learning.”

Well before COVID-19 hit, CDC began an annual competition to most accurately predict the severity and spread of influenza across the United States. The competition receives dozens of entries each year; Biggerstaff says roughly half involve machine learning algorithms, which learn to spot correlations as they are “trained” on vast data sets. Roni Rosenfeld, a computer scientist at Carnegie Mellon University, and colleagues have won the contest five times with algorithms that mine data including Google searches, Twitter posts, Wikipedia page views, and visits to the CDC website.

Teams involved in the flu challenge have now pivoted to COVID-19. They are applying AI in two ways. One aims to spot the first signs of a new disease or outbreak, just as HealthMap did. That requires the algorithms to look for ill-defined signals in a sea of noise, a challenge on which a well-trained human may still hold the upper hand, Pollack says.

AI could also assess the current state of an epidemic—so-called now-casting. The Carnegie Mellon team aims to now-cast COVID-19 across the United States, using data collected through pop-up symptom surveys by Google and Facebook, Google search data, and other sources in order to predict local demand for intensive care beds and ventilators 4 weeks out, Rosenfeld says. “We’re trying to develop a tool for policymakers so that they can fine-tune their social distancing restrictions to not overwhelm their hospital resources.”

Although automated, AI systems are still labor intensive, notes Rozita Dara, a computer scientist at the University of Guelph who has tracked avian influenza and is turning to COVID-19. “By the time you get to AI, it’s the easy part,” she says. To train a program to scan Twitter, for example, re-

searchers must first feed it examples of relevant tweets, painstakingly selected by hand, Dara says. AI may also struggle in a rapidly evolving pandemic, where correlations between online behavior and illness can shift.

AI has misfired before. From 2009 to 2015, Google ran an effort called Google Flu Trends that mined search query data to track the U.S. prevalence of flu. At first the system did well, correctly predicting CDC tallies roughly 2 weeks ahead of time. But from 2011 to 2013, it overestimated flu prevalence, largely because researchers didn’t retrain the system as people’s search behavior evolved, Yom-Tov says. Searches for news reports about the flu were misinterpreted as signs of infection.

“I don’t think it’s an inherent problem,” Yom-Tov adds. “It’s something that we’ve learned from.” He and colleagues from University College London recently posted a paper to the arXiv preprint server showing they could correct for that media-related bias.

Nations struggling to adequately test for the new coronavirus might be tempted to use automated surveillance instead. Biggerstaff says that would be a mistake. When the flu re-emerges this fall, he says, only testing will be able to distinguish outbreaks of it and COVID-19. But AI might help policymakers direct more testing to hot spots. “The hope is that you would actually have the two working together,” says John Brownstein, an epidemiologist at Boston Children’s who co-founded HealthMap in 2006.

Some researchers question whether AI systems will be ready in time to help with COVID-19. “AI will not be as useful for COVID as it is for the next pandemic,” Dara says. Still, machine learning in epidemiology seems here to stay. Pollack, who sounded the alarm about COVID-19 the old-fashioned way, says she, too, is working on an AI program to help scan Twitter for mentions of the disease. ■

SPACE SCIENCE

Crewed launch deepens ties between NASA and SpaceX

Bold company’s rise brings benefits and risks for space science

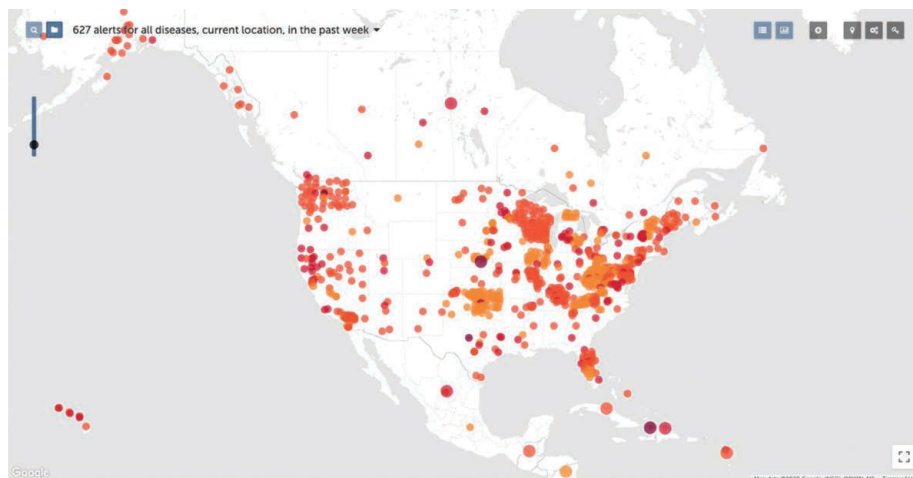
By Adam Mann

On 27 May, NASA will launch people into space from U.S. soil for the first time since 2011, when the space shuttle Atlantis roared aloft on its final voyage. This time, astronauts will be riding to the International Space Station (ISS) not on a NASA rocket, but aboard vehicles bought from the private space company SpaceX: the Dragon 2 capsule atop a Falcon 9 rocket.

The occasion marks yet another milestone for the private California company, which over the past decade has gone from underdog to dominator. SpaceX now handles about two-thirds of NASA’s launches, including many research payloads, with flights as cheap as \$62 million, roughly half the price of a rocket from United Launch Alliance, a competitor. SpaceX’s goals are not limited to low-Earth orbit: Last month it won a contract to build a Moon lander, and it is steadily testing a huge heavy-lift rocket, called Starship, that could carry people to Mars.

Researchers see both benefits and risks in the company’s increasing power. It has lowered the cost of spaceflight through innovations such as reusable stages and fairings, saving NASA money. With its outsize capacity, Starship could cheaply put large telescopes in orbit and heavy science experiments on moons and planets. Yet SpaceX, with a fast-and-loose Silicon Valley mindset, has overlooked the potential for its technologies to contaminate night skies and pristine planets. Some worry the company, led by brazen billionaire Elon Musk, could jeopardize NASA’s long-standing culture of safety. “NASA tries to model everything to the nth degree,” says David Todd, an analyst at Seradata, which tracks launches and satellites. “SpaceX works on the basis of ‘test it until it breaks.’”

Between 2006 and 2008, the first three flights of its Falcon 1 rocket ended in failure.



HealthMap uses artificial intelligence and data mining to spot disease outbreaks (colored dots) and issue alerts.



Next week, SpaceX's Dragon capsule will carry two NASA astronauts into orbit.

SpaceX pivoted to a larger Falcon 9 rocket in 2010, and began to deliver cargo to the ISS for NASA 2 years later. Since then, its ambitions have grown. “A lot of other space companies are trying to win contracts,” says Jonathan McDowell, an astrophysicist at the Harvard-Smithsonian Center for Astrophysics (CfA). “SpaceX is trying to get to Mars. It turns out that having a goal can be economically successful.”

The upcoming crewed flight could displace the Russian rockets NASA has hired—at a hefty price—to carry humans to the ISS since 2011. Cheaper, more frequent flights could improve the biomedical and physical science experiments aboard the station, says industry analyst Laura Forczyk, owner of the space consulting firm Astralytical. “More people equals more research,” she says.

SpaceX has boosted NASA science in other ways, delivering the climate-observing Jason-3 satellite and the planet-seeking Transiting Exoplanet Survey Satellite to orbit. In 2022, it is set to launch the Psyche mission to a metallic asteroid, in the first NASA launch of a Falcon Heavy, which sits between the Falcon 9 and Starship in its propulsive power (*Science*, 26 January 2018, p. 376).

But it's the company's upcoming Starship that has designers of science missions salivating. SpaceX has not announced a date for an inaugural flight, but has built six prototypes at a pace of nearly one per month. (Three have been accidentally destroyed in testing.) The steel alloy capsule and its super-heavy booster stand 120 meters tall, towering over the Saturn V that carried people to the Moon. Last year, Musk said full reusability and thrifty use of propellant would drop

the cost of each Starship launch to \$2 million. Todd suggests \$10 million per flight might be more realistic.

The rocket's 9-meter-diameter cargo hold could easily accommodate giant celestial observatories, such as the proposed Habitable Exoplanet Observatory, which would directly image distant planets. One reason for the endless delays afflicting the James Webb Space Telescope, the successor to the Hubble Space Telescope, has been the need to fold up its segmented 6.5-meter mirror to fit aboard a European Ariane 5 rocket, says CfA astrophysicist Martin Elvis.

A viable Starship could also create political pressure to scupper the Space Launch System (SLS), the NASA-developed heavy-lift rocket that is supposed to power the agency back to the Moon and on to Mars at a whopping \$900 million per launch—if it ever launches. Its debut has repeatedly slipped, and is now expected at the end of 2021. Just 1 or 2 years later, it is supposed to carry astronauts to lunar orbit, but McDowell doubts

it will remain in production for very long. “If Starship works, that's the death knell for SLS,” he says.

SpaceX, along with private companies Blue Origin and Dynetics, was chosen in April to design lunar landers for astronauts and supplies. SpaceX put forth the Starship capsule as its lander, which could launch atop its own booster or a NASA-built one. There would be plenty of room for scientists to piggyback experiments, such as a radio telescope to peer back to the earliest era of galaxy formation from the Moon's far side, says Steve Clarke, NASA's deputy associate administrator for scientific exploration.

Yet SpaceX's haste to go big could also cause trouble for scientists. The Starship lander will be much heavier

than the spindly lunar module of the Apollo missions. The dust and rocks it kicks up could rise into lunar orbit, creating an interfering haze for other landers and threatening satellites and outposts, Elvis says. The company's long-standing goal to colonize Mars has the potential to contaminate the planet with terrestrial microbes that could confound researchers, he adds. SpaceX did not respond to requests for comment for this article.

In recent months, the company has aroused the ire of astronomers with the launch of hundreds of Starlink satellites, which are intended to deliver high-speed internet to remote areas. From the ground, the satellites appear surprisingly bright because of their low orbits, and they have left disruptive trails on the cameras of survey telescopes. “I don't think they intended to screw up people's skies,” says Megan Donahue, president of the American Astronomical Society. “It was just because nobody asked that question of them.”

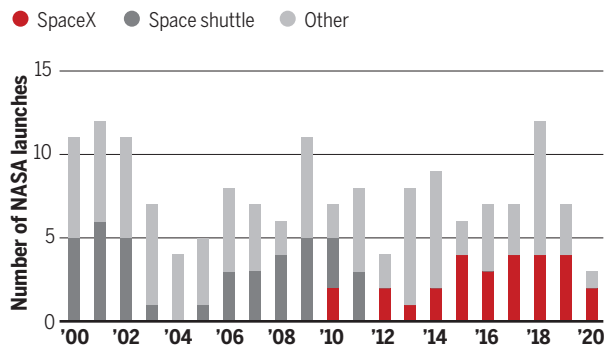
SpaceX is trying to mitigate the issue. Some satellites in the next batch, set to launch soon after the crewed test, will be blackened and equipped with visors that block sunlight. Donahue praises the company for working with researchers to address the problems. “We're all into science,” she says.

But the episode has reminded space scientists not to underestimate SpaceX's potential impacts on their fields. Although Musk is often too bullish about the time scales for his projects, McDowell says, he tends to realize his dreams in the end. “[Musk] has strengths and weaknesses. His overoptimism is kind of both.” ■

Adam Mann is a journalist in Oakland, California.

Going up

Since the retirement of the space shuttle in 2011, SpaceX rockets have picked up an increasing share of NASA's launches.



*Includes launches with a major NASA payload through April, including failures

CLIMATE CHANGE

Tropical forests store carbon despite warming

But if global temperatures reach key threshold, dying trees will release warming gases

By Elizabeth Pennisi

Tropical forests have been one of Earth's best defenses against rising carbon dioxide levels. The trees suck carbon from the atmosphere as they grow, and researchers estimate that, despite ongoing deforestation, tropical forests hold more carbon than humanity has emitted over the past 30 years by burning coal, oil, and natural gas. But scientists have worried that the ability of tropical forests to act as carbon sinks will diminish and ultimately reverse with continued global warming, as trees stressed by heat and drought die and release their carbon.

Now, on p. 869, researchers report that measurements of carbon storage and growing conditions for some 500,000 trees around the world suggest some tropical forests, particularly in Africa and Asia, will—if left intact—continue to sequester large amounts of carbon even as global temperatures rise. But only up to a point. “There are certain levels where forests can’t respond,” says William Anderegg, a forest ecologist at the University of Utah. If warming reaches 2°C above preindustrial levels, the study finds huge swaths of the world’s tropical forests will begin to lose more carbon than they accumulate. Already, the hottest forests in South America have reached that point.

Trees, with their long lives and massive woody trunks, are particularly good at storing carbon. But just how much carbon tropical forests can capture as the planet warms depends on the balance between tree growth spurred by higher atmospheric carbon levels and tree stress and death caused by rising temperatures and increasing drought. “It comes down to a tug of war,” Anderegg says, “between the benefits of carbon dioxide and the potential impact of climate change.”

To see which side might ultimately win, a global team of more than 200 researchers measured more than half a million trees in 813 forests in 24 countries. The team, led by tropical ecologist Oliver Phillips of Leeds University and his postdoc Martin

Sullivan, calculated how much carbon the different forests now store based on the height, diameter, and species of each tree. Researchers also looked at how carbon storage varied from place to place using data from 590 long-term monitoring plots.

To forecast how carbon accumulation might change in the future, the researchers assumed that the hottest forests, which are mostly in South America, are bellwethers of the future. By comparing carbon storage in forests across the range of climates, they could use space as a proxy for time. They analyzed how changes in temperature and precipitation might affect carbon storage, looking for those changes that best explained what they observed in the forests.

ture during the warmest month of the year rises to 32.2°C—long-term carbon storage capacity declines steeply and carbon loss increases. The decline is even greater in drier forests, notes Sullivan, now at Manchester Metropolitan University, likely because the lack of water makes trees more vulnerable to stress and death.

The team calculated that, worldwide, each 1°C increase in maximum temperature reduces carbon storage in tropical forests by 7 billion tons (roughly equivalent to total U.S. carbon emissions over 5 years), although much of that loss is currently offset by increased growth. If global temperatures rise 2°C above preindustrial levels, however, 71% of tropical forests will

be pushed past the thermal tipping point, the researchers found. Carbon losses would be four times greater, with South America seeing the largest loss.

The “massive data compilation ... allows us to draw conclusions with much higher confidence than individual studies would allow,” says Julia Pongratz, a climate scientist at Ludwig Maximilian University of Munich. But ecosystem ecologist Lara Kueppers of the University of California, Berkeley, worries the study might be too optimistic in forecasting that cooler forests, especially in Asia and Africa, will continue to accumulate large quantities of carbon as they warm. It’s not clear whether those forests will

behave like their counterparts in South America, she notes, or that they can adapt to the speed of human-induced climate change. “I don’t have confidence that forests are going to be able to adjust on the time scale they will need to,” she says.

Other researchers see the findings as a wake-up call for action, noting the world has already warmed about 1°C above preindustrial levels. “Even though tropical forest sinks will weaken, conserving them is still better than not having them at all and turning them into carbon sources,” says Richard Betts, a climate modeler specializing in the global carbon cycle at the University of Exeter. “It is not too late,” he adds, “to avoid the most severe impacts.” ■

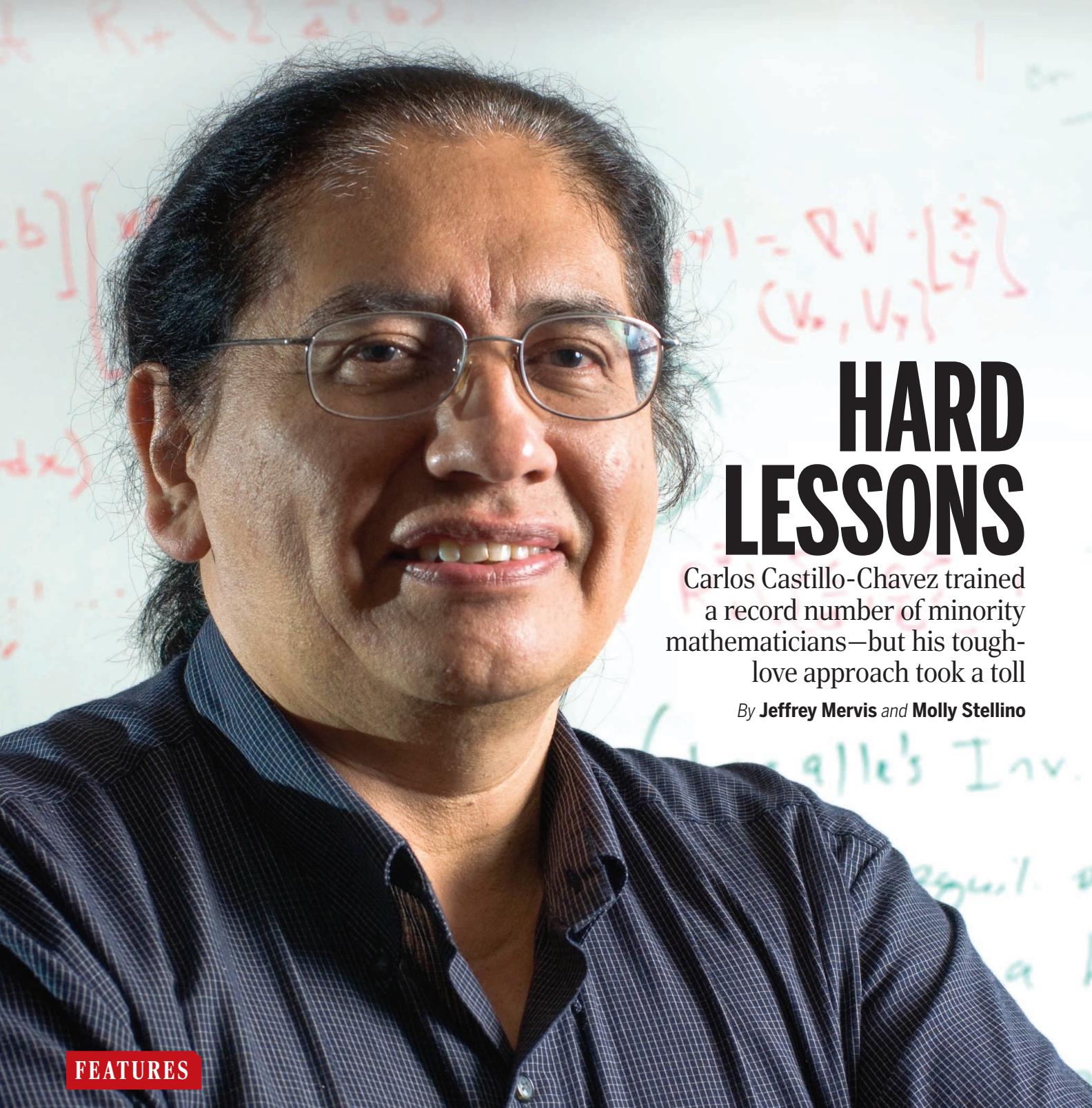


Researchers collect leaves to measure carbon storage in an Andean cloud forest.

(The analysis takes into account differences in the forests’ mix of tree species.)

Previous studies had suggested the lowest temperature a forest experiences at night has the biggest impact on its long-term carbon storage capacity, perhaps because warm nights cause trees to boost respiration and release more carbon. But this study found that the maximum daytime temperature is most important, perhaps because on hot days trees slow their carbon dioxide intake to reduce water loss through pores in their leaves.

The study showed that, overall, the forests now take up more carbon than they lose. But it found that at a tipping point—when the average daily maximum tempera-



HARD LESSONS

Carlos Castillo-Chavez trained a record number of minority mathematicians—but his tough-love approach took a toll

By Jeffrey Mervis and Molly Stellino

FEATURES

The numbers tell the story of a master mentor. Carlos Castillo-Chavez, a Mexican-born mathematical biologist, has trained some 50 Ph.D. students, two-thirds of whom belong to groups historically underrepresented in science. He is especially proud of what he calls his “diamonds in the rough”—students

from less selective undergraduate programs who have ended up with good jobs in academia, industry, and the public sector.

At Arizona State University (ASU), Tempe, where he held the coveted title of regents professor and an endowed chair, Castillo-Chavez presided over a mini-empire of programs designed to increase diversity within the math community. He’s been honored

by three U.S. presidents for expanding the educational horizons of thousands of minority students. His work has been fueled by nearly \$50 million in grants from federal agencies and the private sector.

But that’s all in the past. On 16 May, Castillo-Chavez retired from ASU. It was the final step in the dismantling of his empire.

In August 2019, he abruptly resigned as

director of the Simon A. Levin Mathematical, Computational and Modeling Sciences Center. He also gave up running two programs housed there: a graduate program in applied math and life and social sciences (AMLSS) and the Mathematical and Theoretical Biology Institute (MTBI), a summer research program for undergraduates. A majority of MTBI participants and a significant fraction of AMLSS students are from underrepresented groups. Castillo-Chavez created those programs and ran them with a rare degree of autonomy.

His resignation from those posts, combined with his absence from ASU's Tempe campus since January 2019, have been the cause of much speculation. "I was shocked when I heard," says Ricardo Cordero-Soto, an associate professor of math at California Baptist University, who trained under Castillo-Chavez and who is also active in mentoring minority students. "I thought only death would retire him."

ASU OFFICIALS HAVE SAID nothing about why Castillo-Chavez gave up his positions. "He's gone, and he won't be coming back," was ASU President Michael Crow's terse comment in a 4 March interview with *The State Press*, the university's student-run newspaper. "He's no longer in his historical role."

But *Science* and *The State Press* have learned Castillo-Chavez was stripped of his administrative titles only days after ASU officials resolved a complaint against him by one of his graduate students. The student, Maria Martinez, alleged in her April 2019 complaint that Castillo-Chavez maintained a "hostile work environment," that he "participated in workplace harassment," and that he violated federal laws protecting the rights of persons with disabilities. After a 3-month investigation, ASU officials told Martinez that Castillo-Chavez had agreed to immediately "resign from all administrative appointments at ASU" and would also be retiring.

There was no public announcement, and ASU officials did not release a report or even write up their findings. But in interviews, many colleagues and former students have spoken about what they regard as a dark side to Castillo-Chavez's passion and commitment to diversity: namely, an intolerance for disagreement and a penchant for bullying students and associates.

"Carlos is equally good at building bridges and then burning those bridges," says a mathematician who requested anonymity because of a history of contentious relationships with Castillo-Chavez. "He can work a room and come out with five people who are suddenly his best

friends—until something happens that he doesn't like. And then he turns on them."

"What's sad is that Carlos did some good work," says mathematician Wayne Raskind, a former ASU department chair who left after repeated clashes with Castillo-Chavez. "But he was allowed to go rogue. He eventually became completely full of himself and started to do some bad things. And the more he got away with, the more he did."

Cordero-Soto credits Castillo-Chavez for "helping me realize applied math was the right path for me [and] for looking out for underrepresented students." But Cordero-Soto says he steered one promising student away from ASU because of what he saw as Castillo-Chavez's harsh approach to mentoring graduate students and he believes "compassion" yields better results. "I tell my students that I will be their advocate and their biggest cheerleader because I've been there myself," says Cordero-Soto, who is active in the Math Alliance, a national organization that promotes the mentoring of minority students.

The 68-year-old Castillo-Chavez says he is leaving ASU for reasons that have nothing to do with the complaint. "I'm exhausted and tired," he says. He shed his administrative duties, he says, because he was consumed with caring for his mother, who died in November 2019 after a long illness. His many off-campus commitments were another distraction, he adds.

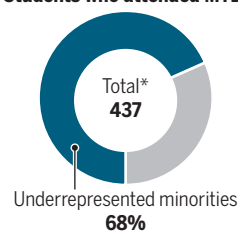
"I've been away from Arizona State for much of the last 7 or 8 years," he says. That period includes a tumultuous 2 years at the helm of a new technical university in Ecuador, 1 year of cancer treatment in Boston, and, most recently, a visiting professorship at Brown University. "I've been trying to handle things from far away, and I'm burned out."

He says he had planned to retire in December 2020 and simply moved up the date. But he acknowledges he could have done a

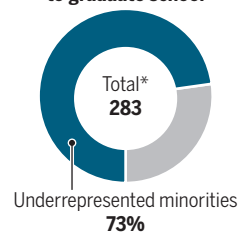
The institute that could

Carlos Castillo-Chavez's undergraduate research program, the Mathematical and Theoretical Biology Institute (MTBI), has boosted math diversity since 1996.

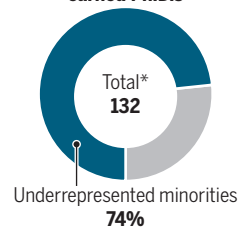
Students who attended MTBI



MTBI participants who went to graduate school



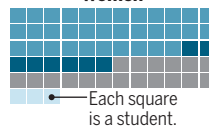
MTBI participants who have earned Ph.D.s



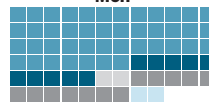
Demographics of Ph.D.s



Women



Men



*Totals represent U.S. citizens or permanent residents.

better job of addressing the issues Martinez raised.

"I think her complaint had some validity," he says. "I did not keep track of the potential seriousness of the situation. I was responsible for her [training], and I accept that responsibility."

The past year marks a somber end to an improbable journey that took Castillo-Chavez to the pinnacle of his profession. Born in 1952 to a working-class family in Mexico City and radicalized by the failed 1968 pro-democracy protests there, Castillo-Chavez hoped to use community theater as a vehicle for social activism. But he pivoted to academia after losing a student acting contest.

Emigrating to the United States in 1974, he quickly earned a bachelor's degree and started a Ph.D. program in math at the University of Wisconsin (UW), Milwaukee. However, a dinner conversation with his adviser and other faculty members nearly derailed his plans. The professors talked disparagingly about a university memo that described efforts to attract more minority students, he says, adding that it was clear to him they thought a Latino student wasn't capable of making it on their own. "So, I quit," he told the *Lathisms* podcast in 2018.

After reconsidering, he resumed graduate school at UW's flagship campus in Madison and earned his Ph.D. in 1984. A postdoc at Cornell University led to a faculty position—and, eventually, a tenured professorship—in what was then the biometrics department.

Hoping to launch others on a similar path, Castillo-Chavez founded MTBI in 1996. The 8-week summer program combines graduate-level courses with a group research project that targets a real-world problem. "What I have done over the last 20-plus years is to take students from nonselective schools and show them that their school of origin is irrelevant," Castillo-Chavez told *Lathisms*.

Of the roughly 530 undergraduates from around the world who have attended MTBI (see graphic, p. 815), more than 160 have gone on to earn a Ph.D., he says—nearly 60% of whom are from groups historically underrepresented in science. Those graduates represent roughly 40% of all Latino Ph.D.s in math from U.S. universities in the past decade, he says.

“Carlos Castillo-Chavez has been a giant in the field of STEM [science, technology, engineering, and math] education with respect to underrepresented minorities over a period of almost 2 decades,” Ted Greenwood, a former program officer at the Alfred P. Sloan Foundation, which funded MTBI for many years, wrote in a 60th birthday tribute to him.

MTBI’s reach extended far beyond ASU. “Carlos began sending us alumni of his MTBI program, most of them Latino,” recalled Phil Kutzko, an emeritus professor of math at the University of Iowa. “Without his trust and support, it is unlikely that the transformation of our graduate program ... would have occurred,” Kutzko wrote in a recent essay about the Math Alliance, which he co-led for many years.

Carlos Torre, who earned his Ph.D. under Castillo-Chavez in 2009 and has spent the past decade on Wall Street, fell under his spell as a Cornell undergraduate. “It was the first time I had a professor who looked and sounded like me,” recalls Torre, who as an infant immigrated with his family from Peru. “He had a very thick accent, and a ponytail.”

Torre was intrigued when Castillo-Chavez mentioned MTBI one day in class. He lobbied for a slot, and over the next several years he received several doses of Castillo-Chavez’s tough-love medicine.

“He called me out when I should have been called out,” Torre says, recalling a tongue-lashing he once received after dropping a difficult graduate course that wasn’t required for his degree. As punishment, Castillo-Chavez barred Torre from doing his own research that summer at MTBI.

Torre wasn’t happy, but he thinks Castillo-Chavez was right. “It also meant a lot to me that somebody cared so much.”

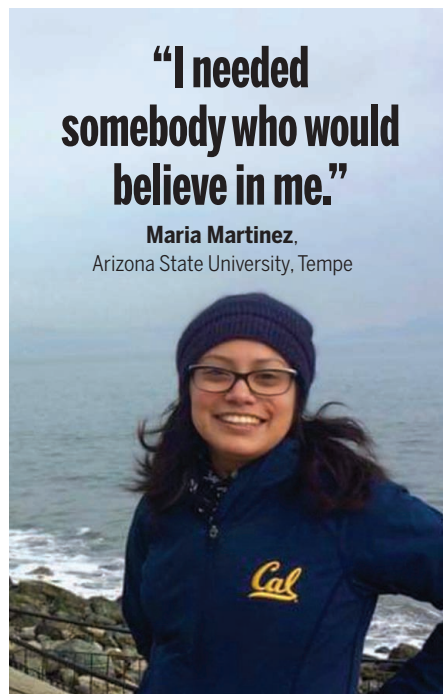
CASTILLO-CHAVEZ WAS RECRUITED to ASU by Joaquin Bustoz, a beloved figure and former chair of the math department who died in a car crash before Castillo-Chavez arrived. Castillo-Chavez immediately took over a summer program for minority high school students that Bustoz had run for almost 2 decades, adding it to the MTBI program as part of his portfolio for increasing diversity.

The next step, he decided, was a graduate program in applied math. But

Castillo-Chavez says he failed to win over his colleagues in the math department.

“Carlos had created an enormous amount of friction” in trying to obtain more resources, recalls one longtime member of the department, who requested anonymity. “And the dean decided that could not go on. So Carlos was given the opportunity to start his own program [in a different department]. There were not many people in the department who regretted seeing him leave.”

The result was the Levin center, located in ASU’s School of Human Evolution and Social



**“I needed
somebody who would
believe in me.”**

Maria Martinez,
Arizona State University, Tempe

Change. It serves as the administrative home for AMLSS, a graduate program for applying math to model how social forces affect the origin and spread of disease. Named to honor Simon Levin, a prominent theoretical biologist and Castillo-Chavez’s mentor at Cornell, the center doesn’t have its own faculty. Instead, its work is carried out by visiting scholars drawn from Castillo-Chavez’s vast network of contacts.

Even as Castillo-Chavez’s empire grew and accolades rolled in, so did conflicts with colleagues and students. Former colleagues say senior ASU officials either ignored reports of questionable behavior or defended him because his successful track record burished the university’s reputation.

Raskind says Castillo-Chavez maligned him in emails to top ASU officials, including Crow, and threatened to destroy his career after Raskind put in a bid for a deanship. (Raskind would later serve as dean at two other research universities.)

Raskind says his offense was saying no to a request that his department provide computing resources for the new center.

Senior officials ignored Raskind’s complaints about Castillo-Chavez’s behavior, he says, “and that’s when I knew it was time to leave. I was a tenured professor, so he couldn’t ruin my career. But he tried to do similar things to some of his students, which I think was disgraceful.”

Other former and current colleagues tell similar stories. “Where it became toxic is that, whenever you would get into a disagreement with Carlos, he would make it an issue about minorities,” says one ASU faculty member. “He would send emails to everybody at the university, insinuating that you were a racist.”

One former student, who says he never saw Castillo-Chavez act maliciously, thinks professional jealousy may have been a factor in some of those clashes. “The fifth floor [of one ASU building] was all Carlos,” says Daniel Rios-Dorio, who joined AMLSS in 2006 after participating in MTBI as a graduating senior from the City University of New York. As many as 20 students joined the program every fall, and each got a laptop, plenty of space to work, and travel funds to attend conferences. “The man was clearly doing something right,” says Rios-Dorio, who earned his Ph.D. in 2010.

AMLSS has no prescribed set of courses and no qualifying exams. Castillo-Chavez thought such requirements might penalize students with weaker backgrounds and believed a student’s research and publications were better metrics. But that flexibility also allowed Castillo-Chavez to make random and arbitrary demands, such as suddenly announcing that students must attend a new course being offered by a visiting colleague.

“He had different requirements for different students, and nobody knew what any of those requirements were,” says Jack Pringle, a second-year graduate student in the program. “As soon as I arrived [in 2018], it was very clear that everyone was terrified of Carlos. But there was also this worship of Carlos. And I think he liked having that power and dependency.”

Castillo-Chavez says such comments, which he says he’s heard for years, mischaracterize his approach to mentoring.

“I think that some students did not understand the rigors of the program, and that its flexibility created challenges that some students were unable to meet,” he says. “I have devoted my life, and sacrificed my family, to this cause. But if you want to stand up and promote diversity and criticize the lack of diversity [at an institution], there will be people who accuse you of

wanting to lower academic standards and dilute the quality of the program.”

ONE FORMER STUDENT, Anarina Murillo, says Castillo-Chavez gave her the courage to pursue a career she thought was beyond her grasp.

“Carlos could tell that I suffered from imposter syndrome,” Murillo says, referring to the feeling of many minority students that they don’t belong and that their shortcomings will be exposed. “So, he told me to look in the mirror and say, ‘I’m going to get a math Ph.D.’ He told me to say it over and over until I believed it.”

Murillo was a junior at ASU when she applied for and was accepted into MTBI. “And that’s when I fell in love with mathematical modeling,” she says.

Her academic career took off from there. She earned a Bachelor of Science in 2010 and then ripped through the AMLSS graduate program. After doing a postdoc at the University of Alabama, Birmingham, she accepted Castillo-Chavez’s invitation last year to join him at Brown University. Within months she was hired as an assistant professor in Brown’s School of Public Health, where she applies her math background to the design of clinical pediatric studies.

“I always knew that I wanted to be a professor,” she says. “But I never expected it to happen.”

Maria Martinez hoped Castillo-Chavez might do the same thing for her academic career, which was in crisis.

In 2011, Martinez was sexually assaulted shortly after starting a Ph.D. program at the University of California (UC), Berkeley. The injuries she suffered eventually required five surgeries and the use of a wheelchair for 18 months. The experience also led her to rethink her decision to pursue a degree in theoretical math.

Martinez says she was moved by Castillo-Chavez’s passion for helping minority students when she heard him speak at a Latinx in the Mathematics Sciences conference in Los Angeles in March 2018. And she thought AMLSS was an ideal setting to pursue her newfound interest in the evolution of societal attitudes toward violence against women.

“This is exactly what I needed,” Martinez remembers saying to herself. “The whole idea behind his program is to apply mathematics to problems in the social sciences. I also needed somebody who would believe in me.”

Martinez says Castillo-Chavez promised her generous funding and “a safe space to learn and work.” But her hopes for a brighter future faded quickly after she arrived in August 2018.

In one case, she says, Castillo-Chavez “be-

came very agitated” and berated her for talking to a reporter who wanted to write about her experience at UC Berkeley. (No story ever appeared.) He repeatedly cited her ongoing medical issues in conversations about her academic status, she adds. Castillo-Chavez also threatened to renege on a promise to bring her to Brown, she says, “because you’ve been sick so often.”

Martinez says Castillo-Chavez questioned her commitment to the program and her ability to complete her degree. After denying her request to miss 2 days to attend her brother’s medical school

dean of the College of Liberal Arts and Sciences. “No decisions have been made regarding how we will move forward next year and beyond,” Kenney said earlier this month.

Castillo-Chavez, who has applied for emeritus status, will not have a say in those decisions. But he is worried about what might happen.

“The future of everything I have created is uncertain,” he says. “I’m not sure there’s anybody else who would be willing to do what I’ve had to do to keep it going.” As one faculty member who has clashed with him puts it, “a center without Carlos is not some-



Carlos Castillo-Chavez (center) with students from his undergraduate summer research program.

graduation, she says he told her she was “a bad student, ... not good at research, ... undisciplined, and a troublemaker.” In her 23 April 2019 complaint to Crow, Martinez wrote, “Dr. Castillo ... created an environment so tense that I am scared to say anything to anyone.”

ASU OFFICIALS DECLINED to discuss any aspect of their investigation. But emails Martinez provided describe how the university’s associate general counsel, Becky Herbst, sought her approval for an “informal” resolution of her complaint in which Castillo-Chavez would agree to step down from his leadership positions and retire. The findings of the investigation were never made public. Martinez accepted that solution, she says, because “he could have hurt so many other people if he remained.”

The fate of the empire Castillo-Chavez built is “under review,” says Patrick Kenney,

thing that anybody has ever considered.”

The future is equally murky for Martinez, who last year moved back to her hometown of Los Angeles after finding another adviser from a different ASU department. “I’d like to teach math at a community college in the Los Angeles area if campuses reopen in the fall,” she says, noting that she’ll be taking a leave of absence at the end of this semester to weigh her options. It would be a way to give others the chance at a quality education that her parents, who were factory workers, were denied.

However, she doesn’t know whether she’ll ever finish her dissertation and earn a Ph.D. “I want to be a mathematician,” she says. “But after my experience with Dr. Castillo-Chavez, I don’t know if I can.” ■

Molly Stellino graduated from Arizona State University, Phoenix, this month and served as editor of *The State Press Magazine*.

INSIGHTS

PERSPECTIVES

A Bangladeshi woman collects potable water from a hand pump. The health toll from arsenic in water justifies global solutions.

PUBLIC HEALTH

Global solutions to a silent poison

Modeling arsenic in domestic well water highlights large data gaps in testing

By Yan Zheng

Severe public health consequences of worldwide geogenic arsenic occurrence in groundwater have been recognized since the late 1990s (1). The population affected by groundwater arsenic from domestic well supplies has been frequently stated to exceed 100 million. However, this compilation is fraught with uncertainties due to incomplete and unreliable records on domestic wells that supply drinking water and incomplete testing for arsenic. On page 845 of this issue, Podgorski and Berg use statistical models to estimate that 94 million to 220 million people, with 94% in Asia, are at risk of

drinking well water containing arsenic concentrations >10 $\mu\text{g}/\text{liter}$ (2). In Bangladesh, a 2009 national drinking-water quality survey found that about 20 million and 45 million people were exposed to concentrations greater than 50 and 10 $\mu\text{g}/\text{liter}$, respectively, with an arsenic-related mortality rate of 1 in every 18 adult deaths (3). This global threat demands multisector solutions.

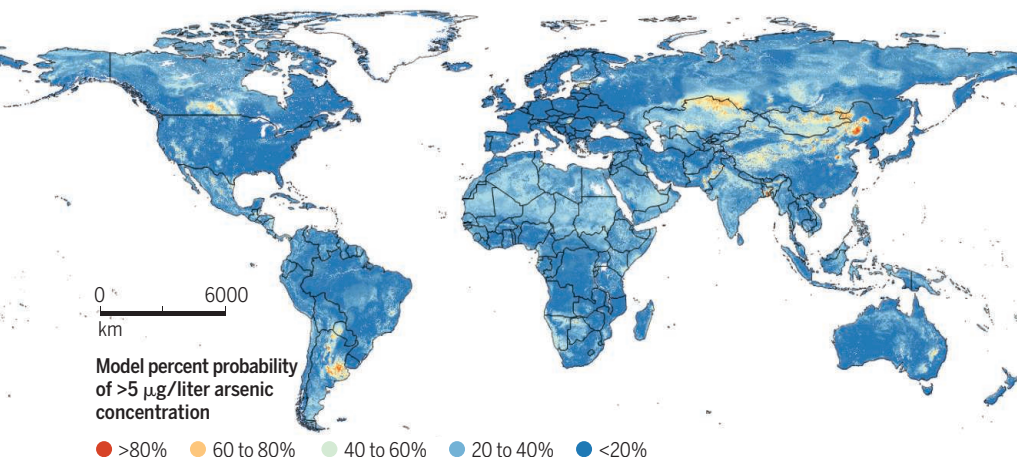
The 2017 edition of the World Health Organization's *Guidelines for Drinking-Water Quality* retained its provisional value of 10 $\mu\text{g}/\text{liter}$ for inorganic arsenic, a recommendation based on treatment performance and analytical achievability. Many countries have adopted this value as their drinking-water quality standard over the past two decades. Although the European Union has set the standard at 10 $\mu\text{g}/\text{liter}$, Denmark's is more protective at 5 $\mu\text{g}/\text{liter}$. The Association of Dutch Drinking Water Companies

voluntarily agreed on a guideline of <1 $\mu\text{g}/\text{liter}$ in 2015 (4). The U.S. Environmental Protection Agency adopted 10 $\mu\text{g}/\text{liter}$ in 2001 for the federal maximum contaminant level (MCL) on the basis of cost-benefit analyses, effective 2006. However, the state of New Jersey opted for 5 $\mu\text{g}/\text{liter}$ effective 2006, and New Hampshire adopted 5 $\mu\text{g}/\text{liter}$ in 2020. The world's two most populous countries, China and India, lowered their MCL, effective 2007 and 2012, respectively, from 50 to 10 $\mu\text{g}/\text{liter}$. However, 50 $\mu\text{g}/\text{liter}$ is permissible in the absence of alternate sources in India and remains the MCL for Bangladesh and for small, dispersed rural supplies in China. This order-of-magnitude disparity in the MCL is concerning because new health evidence suggests that even 10 $\mu\text{g}/\text{liter}$ may not be protective enough, especially during early, biologically vulnerable stages of life.

School of Environmental Science and Engineering,
Southern University of Science and Technology, Shenzhen
518055, China. Email: yan.zheng@sustech.edu.cn

A world model for groundwater arsenic risk

Lowering arsenic concentrations in drinking water helps avoid a range of adverse health outcomes. Modeling the probability of groundwater arsenic with excess risks helps guide testing. Podgorski and Berg developed global models for groundwater arsenic concentrations exceeding 5 and 10 $\mu\text{g}/\text{liter}$.



The word arsenic originates from the Greek *arsenikon*, which means valiant, bold, or potent. Odorless and tasteless when dissolved in water, this silent poison became known as both “the king of poisons” and “the poison of kings.” The acute toxicity of inorganic arsenic, classified as a group I carcinogen by the International Agency for Research on Cancer, has been appreciated since ancient times. Long-term exposure to water containing high concentrations ($>100 \mu\text{g}/\text{liter}$) of inorganic arsenic (arsenate and arsenite) is associated with nonmelanoma skin, lung, and bladder cancers, as well as noncancer outcomes. The Health Effects of Arsenic Longitudinal Study (HEALS) in Bangladesh showed dose-response relationships between drinking-water arsenic and skin lesions, respiratory symptoms, cardiovascular disease, and reduced intellectual function in children (5). Long-term exposure to moderate concentrations ($<50 \mu\text{g}/\text{liter}$) has been associated with cardiovascular disease incidence and mortality in one of the largest studies in the United States (6). Epidemiologic evidence, consistent with experimental evidence, supports that arsenic affects birth outcomes and impairs neurodevelopment when exposure occurs during early life, even at moderate concentrations ($<50 \mu\text{g}/\text{liter}$) (5). In utero, arsenic exposure has been associated with alterations in gene expression pathways related to diabetes (7), which may contribute to adult diabetes risks. This supports the epigenome as a general mechanism involved in arsenic toxicity, consistent with evidence from a genome-wide DNA methylation study of 396 HEALS adults (8). Still, not enough is known about the mode of action of inorganic arsenic for extrapolating dose response to very low concentrations ($<5 \mu\text{g}/\text{liter}$).

Because three-dimensional (longitude, latitude, and depth) mapping of groundwater arsenic concentration often lacks the spatial resolution to characterize most aquifers, exposure assessment has turned to “predictive” models incorporating geo-environmental predictor variables. Podgorski and Berg utilized 58,555 aggregated well ($<100\text{-m}$ depth) water arsenic average values, mapped to 1-km^2 grid cells based on $>200,000$ tests from 67 countries, to develop a random forest machine-learning model to globally quantify exposed populations. This represents a culmination of logistic regression (9, 10) and machine-learning (11) modeling efforts (see the figure). The authors’ efforts expose data gaps because few countries have conducted a nationwide groundwater arsenic survey. Testing data are also clustered with uneven and incomplete spatial coverage. More arsenic data and detailed predictor datasets will reduce the large and partially unknown uncertainties. Eleven out of 52 spatially continuous predictor variables representing various climatic, geologic, soil, and other parameters emerged through recursive feature elimination to create the simplest best model. Additional research is required to explain why these are important. Statistical models are not meant to predict individual well water arsenic concentrations. Their greatest value lies in identifying potential areas at risk that have not had testing.

This public health crisis leads to an urgent call to test all domestic well water for arsenic worldwide. Testing should prioritize the high-risk areas identified by models. Heterogeneous groundwater arsenic spatial distribution (10^1 to 10^3 m) should make wells that are close to known high-arsenic wells testing priorities. The combination of arsenic’s toxicity and its wide distribution makes this task imperative. Disparities in

Health effects in adults

General health effects

Mortality
DNA methylation
Gene expression

Nervous system

Movement and motor function
Neuropathy

Immune system

Infections

Respiratory system

Bronchiectasis
Lung cancer

Cardiovascular system

Heart and vascular disease
High blood pressure
Stroke

Endocrine system

Diabetes

Soft organs

Kidney cancer
Bladder cancer
Liver cancer

Skin

Skin lesions
Skin cancer

Health effects in children

General health effects

Infant mortality
Reduced birth weight
DNA methylation
Gene expression

Nervous system

Neurological impairment

coverage of regulatory requirements in the United States have left more than a million rural Americans unknowingly exposed to arsenic, with a high proportion belonging to socioeconomically and behaviorally vulnerable groups (10, 12). Development of sensitive, reliable, inexpensive, and user-friendly testing methods for inorganic arsenic in water and urine, preferably with on-site rapid measurement capability, can further improve screening and identify exposed populations. Whereas many countries have succeeded in replacing noncompliant arsenic domestic wells with alternative supplies or treatment to reduce exposure, dispersed rural populations require sustained attention. Treatment of arsenic is not cheap, burdening rural households even in high-income countries. Geogenic arsenic in well water is forever, but our exposure to it should not be. ■

REFERENCES AND NOTES

1. D. K. Nordstrom, *Science* **296**, 2143 (2002).
2. J. Podgorski, M. Berg, *Science* **368**, 845 (2020).
3. S. V. Flanagan, R. B. Johnston, Y. Zheng, *Bull. World Health Organ.* **90**, 839 (2012).
4. A. Ahmad et al., *Environ. Int.* **134**, 105253 (2020).
5. National Research Council, “Critical aspects of EPA’s IRIS assessment of inorganic arsenic: Interim report” (The National Academies Press, Washington, DC, 2013).
6. K. A. Moon et al., *Ann. Intern. Med.* **159**, 649 (2013).
7. A. Navas-Acien et al., *Curr. Diab. Rep.* **19**, 147 (2019).
8. K. Demanelis et al., *Environ. Health Perspect.* **127**, 057011 (2019).
9. L. Winkel et al., *Nat. Geosci.* **1**, 536 (2008).
10. J. D. Ayotte et al., *Environ. Sci. Technol.* **51**, 12443 (2017).
11. J. D. Ayotte et al., *Environ. Sci. Technol.* **50**, 7555 (2016).
12. Y. Zheng, S. V. Flanagan, *Environ. Health Perspect.* **125**, 085002 (2017).

ACKNOWLEDGMENTS

J. D. Ayotte, A. Navas-Acien, and D. K. Nordstrom provided comments. Figure courtesy of A. Bozack, Z. Tan, and B. Xu. Y.Z. is supported by Strategic Priority Research Program of the Chinese Academy of Sciences (XDA20060402), the National Natural Science Foundation (41831279), and the U.S. National Institute of Environmental Health Sciences, Superfund Research Program (P42 ES010349).

10.1126/science.abb9746

CHEMICAL PHYSICS

Tracking both ultrafast electrons and nuclei

Electron diffraction correlates the excited-state decay of pyridine with its ring distortion

By Wolfgang Domcke¹ and
Andrzej L. Sobolewski²

Excitation of polyatomic molecules with visible or ultraviolet (UV) light to a higher-energy electronic state results in a complex competition between radiative and radiationless electronic decay processes and photochemical reactions. Although the time evolution of the population probability of excited electronic states has been extensively explored with time-resolved laser spectroscopy in recent decades, the accompanying nuclear motion could so far not be resolved at the fastest (femtosecond) time scales. On page 885 of this issue, Yang *et al.* (1) report the simultaneous experimental detection of the excited-state decay and the associated deformation of the nuclear frame with subpicosecond time resolution and subnanometer structural resolution for the example of the pyridine molecule.

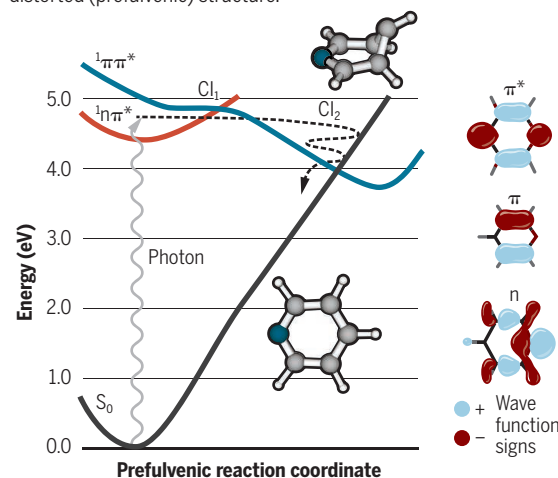
The simplest framework for describing the motion of nuclei and electrons in a molecule is the Born-Oppenheimer (BO) approximation, which assumes that the much faster electronic motions can be calculated separately from the much slower nuclear motions. Traditionally, radiationless transitions in polyatomic molecules were described in a theoretical framework that assumes weak deviations of the nuclear motion from the BO approximation. This concept was appropriate for decay time scales on the order of nanoseconds, but femtosecond time-resolved laser spectroscopy provided ample evidence that radiationless transitions can occur on ultrafast time scales (just tens of femtoseconds) that approach the periods of high-frequency vibrations. This evidence required a profound conceptual revision of the description of radiationless transitions.

The current understanding of ultrafast radiationless transitions is that they are driven by so-called conical intersections (CIs), which are manifolds of exact

degeneracy of electronic potential-energy surfaces at which divergent non-BO couplings cause a complete breakdown of the BO approximation (2). For example, the first excited state and the ground state of a molecule have the same energy at a CI of these states, but the nuclear motion will be subject to different forces. Ultrafast excited-state deactivation through CIs plays an essential role in the protection of fundamental biological molecules (such as DNA and proteins) from photodamage by UV radiation (3).

Photophysics of pyridine

Yang *et al.* used ultrafast electron diffraction to reveal both the structural and electronic changes that occur when photoexcited pyridine relaxes back to the ground state. The excited state has a distorted (prefulvenic) structure.



Excitation and decay pathway

Pyridine in its electronic ground state (S_0) is photoexcited into the $n\pi^*$ (red) excited state. Its relaxation pathway bypasses conical intersection 1 (CI_1) to the $\pi\pi^*$ (blue) excited state toward the prefulvenic structure. At CI_2 , the transition back to the S_0 state occurs.

The photochemistry of benzene and of aza-arenes such as pyridine played a paradigmatic role for the understanding of radiationless transitions through the discovery, in the early 1970s, of the so-called “channel-three” phenomenon, which is a sudden increase in the radiationless decay rate at a certain excess energy in the lowest excited singlet (S_1) state (4). Extensive spectroscopic studies attributed the channel-three effect to an abrupt onset of intrastate vibrational relaxation (IVR) and a sudden shortening of the lifetime of vibrational

levels (5). The generic mechanism behind the channel-three effect was revealed by early *ab initio* calculations for benzene (6, 7). It involves a low-barrier reaction path to a biradical structure which was termed “prefulvene” (8) because it is geometrically related to the valence isomer fulvene. Along the reaction path to prefulvene, a low-lying CI exists at which ultrafast decay from the S_1 energy surface to the energy surface of the electronic ground state (S_0) can occur.

In pyridine, the lowest singlet excited state is of $n\pi^*$ character, where n denotes a “nonbonding” orbital mainly localized on the nitrogen atom and π^* denotes the lowest unoccupied orbital that is delocalized over the six-membered ring. Qualitative potential-energy profiles along the reaction path to the prefulvenic form of pyridine are displayed in the figure. The energy profile of the $1n\pi^*$ state is crossed by the energy profile of the $1\pi\pi^*$ state at a CI marked CI_1 . Beyond a plateau, the $1\pi\pi^*$ energy in turn crosses the energy profile of the S_0 state at the CI marked CI_2 . This specific model proposed in the 1990s established a direct relation between a photophysical phenomenon (radiationless decay) and a photochemical reaction (photoisomerization to fulvene) (9).

Although this general scenario of ultrafast radiationless decay is now widely accepted, it has not been confirmed so far by direct experimental observation. Time-resolved population probabilities of electronic states can now be measured routinely by femtosecond laser spectroscopy with a variety of detection schemes. Time scales of ultrafast radiationless transitions have been established for numerous molecular systems, but these

measurements do not provide information on the nuclear motion driving the electronic transition. Molecular structure can be determined with diffraction methods, and electron diffraction (ED) can be applied to gas-phase samples. In the 1990s, Zewail and co-workers pioneered the development of nonstationary (time-resolved) ED (10). A femtosecond UV pump pulse excites the molecular sample, and diffraction of a time-delayed electron pulse provides structural information. However, the time resolution was limited to ~ 10 ps.

¹Department of Chemistry of the Technical University of Munich, Munich, Germany. ²Institute of Physics of the Polish Academy of Sciences in Warsaw, Warsaw, Poland. Email: domcke@ch.tum.de

The recent development of brilliant pulsed x-ray sources has generated new opportunities for time-resolved diffraction experiments in the femtosecond regime. The experiment of Yang *et al.* is an ED experiment performed at the SLAC-MeV-UED facility (17) on a target gas of pyridine. The pump laser launches a nonstationary wave packet on the potential-energy surface of the $S_1(n\pi^*)$ excited state. Large-angle elastic scattering encodes information on the nuclear structure, whereas small-angle inelastic scattering is sensitive to electron correlation. In the electronic ground state of pyridine, the localized n orbital is doubly occupied, which results in strong so-called dynamical electron correlation (the two electrons try to avoid each other). In the $S_1(n\pi^*)$ excited state, these two electrons occupy spatially separated orbitals, which reduces dynamical electron correlation. The population of the $S_1(n\pi^*)$ state can be detected by the reduction of the small-angle inelastic electron scattering signal in the excited state.

Yang *et al.* extracted the nuclear structural dynamics from the simultaneously measured large-angle elastic ED, using algorithms that were developed earlier for stationary ED. Specifically, the main geometric parameters are the average bond length of the C_5N ring and the dihedral angle representing the distortion of one of the atoms out of the plane of the six-membered ring. The transient structure confirms the prefulvenic distortion predicted earlier by *ab initio* calculations (see the figure). The time-resolved ED data unequivocally reveal that the decay of the population of the $S_1(n\pi^*)$ state and the distortion of the ring occur on the same time scale of ~ 300 fs, resolving a decades-old puzzle in molecular spectroscopy. This work of Yang *et al.* represents a milestone on the path toward the characterization of photochemical events with simultaneous complete resolution in time as well as in atomic structure. ■

REFERENCES AND NOTES

1. J. Yang *et al.*, *Science* **368**, 885 (2020).
2. D. R. Yarkony, in *Conical Intersections: Electronic Structure, Dynamics and Spectroscopy*, W. Domcke, D. R. Yarkony, H. Köppel, Eds. (World Scientific, 2004), chap. 2.
3. S. Boldissar, M. S. de Vries, *Phys. Chem. Chem. Phys.* **20**, 9701 (2018).
4. J. H. Callomon, J. E. Parkin, R. Lopez-Delgado, *Chem. Phys. Lett.* **13**, 125 (1972).
5. E. Riedle, T. Knittel, T. Weber, H. J. Neusser, *J. Chem. Phys.* **91**, 4555 (1989).
6. S. Kato, *J. Chem. Phys.* **88**, 3045 (1988).
7. A. L. Sobolewski, C. Woywod, W. Domcke, *J. Chem. Phys.* **98**, 5627 (1993).
8. D. Bryce-Smith, H. C. Longuet-Higgins, *Chem. Commun.* **1966**, 593 (1966).
9. I. J. Palmer, I. N. Ragazos, F. Bernardi, M. Olivucci, M. A. Robb, *J. Am. Chem. Soc.* **115**, 673 (1993).
10. V. A. Lobastov *et al.*, *J. Phys. Chem. A* **105**, 11159 (2001).
11. X. Shen *et al.*, *Struct. Dyn.* **6**, 054305 (2019).

TOPOLOGICAL OPTICS

Solitons and topological waves

A laser-fabricated waveguide array creates a nonlinear medium that supports solitons

By Mark J. Ablowitz and Justin T. Cole

The intense coherent emission from lasers enabled the study of light propagation in nonlinear media, which spurred many important applications. More recently, the study of electromagnetic wave propagation in periodic media, where linear band structures play an important role, has advanced in new directions. By breaking certain symmetries, such as time reversal, the medium can support so-called “topologically protected” modes that possess uncommon robustness to material defects. Theory has suggested that certain nonlinear waves can inherit the topology of associated linear waves. On page 856 of this issue, Mukherjee and Rechtsman (1) describe experiments where such nonlinear waves, called solitons, can now be observed in the bulk of photonic topological media. These localized waves exhibit cyclotronic motion as the light propagates down a specifically engineered waveguide. When a different mode is considered—one with trivial topology—the waves no longer circulate but remain essentially fixed in their initial spatial distribution.

Department of Applied Mathematics, University of Colorado, Boulder, CO 80309, USA. Email: mark.ablowitz@colorado.edu; justin.t.cole@colorado.edu

“Theory has suggested that certain nonlinear waves can inherit the topology of associated linear waves.”

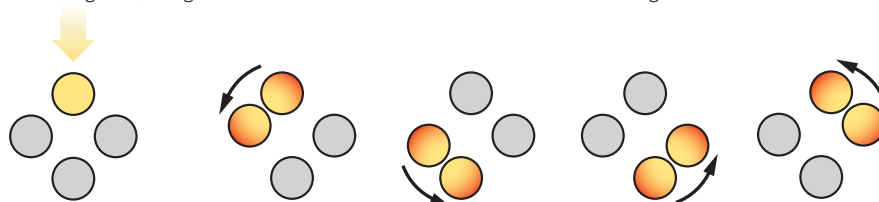
Investigations of solitons trace their roots back to 1834, when naval architect John Scott Russell first recognized their remarkable character in the Union Canal near Edinburgh, Scotland (2). This wave was not oscillatory; it was a solitary surface wave that propagated over surprisingly long distances (2 to 3 km) with fixed form. Some years later, mathematicians described this solitary wave in terms of approximate equations derived from the governing water-wave equations. For nearly

70 years, this was essentially all that was known theoretically. The situation changed in 1965 (3) when it was found that two such solitary waves have extraordinary interaction properties. Their interaction is elastic in nature, and the two waves exit the interaction with the same amplitude and speed with which they entered. Such solitary waves were termed solitons.

This paper motivated major research studies in both mathematics and physics. In mathematics, it gave rise to a new field of study: integrable nonlinear wave systems. These are nonlinear partial differential equations that are exactly solvable and possess an infinite number of symmetries and conservation laws. In physics, researchers have observed solitary waves and solitons not only in water waves and nonlinear optics but also in plasmas, electrical circuits,

Tracking a topological soliton

Five snapshots (left to right) show a topological soliton as it propagates down an array of waveguides laser-fabricated by Mukherjee and Rechtsman. The waveguide is fabricated in such a way that as adjacent waveguides come close together, the light transfers in a counterclockwise fashion from one waveguide to the next.



Light injection

The system is initialized by injecting light (yellow) into a waveguide (shown as gray without light).

Rotating solitons

Light propagates down the waveguide and cyclically returns to its initial state. If the mode was nontopological, it would start and remain in a state like the leftmost figure.

and Bose-Einstein condensates. These solitary waves do not necessarily have special interaction properties.

In a different research direction, it was shown (4) that topological properties and invariants could be used to explain the integer quantum Hall effect. As an external magnetic field is gradually increased, the conductivity in materials such as gallium arsenide heterostructures decreases by quantized jumps. The field of topological insulators in electromagnetic media can be traced to 2008, when topologically protected modes were theoretically identified in suitably constructed material permittivity (5). By using media that break time-reversal symmetry, linear edge waves were found that propagate unidirectionally and possess nontrivial topological invariants (Chern numbers).

The first experimental observation of a topologically protected mode in an electromagnetic system found that localized edge waves could propagate in suitable magnetic media that break time-reversal symmetry (6). These linear magneto-optical waves propagated unidirectionally without backscatter from defects. The topology here is spectral in nature and is different from spatial topology observed in dark solitons and vortices.

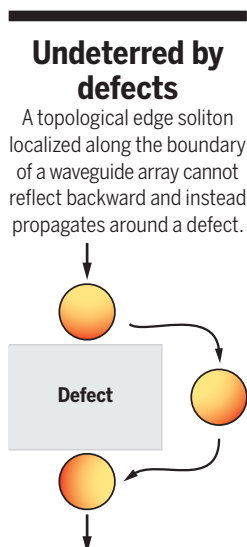
A few years later, a photonic system was constructed (7) that broke time-reversal symmetry by creating a helical rotation of the lattice waveguides in the propagation direction. The media vary periodically in the direction of propagation, and models of this system involve wave equations with coefficients that share this periodicity. The mathematician Gaston Floquet studied differential equations with periodic coefficients, so this system is referred to as a Floquet topological insulator.

All of these systems are linear. A nonlinear waveguide system was proposed in (8) that exhibits a similar type of cyclo-tronic motion, as has now been observed by Mukherjee and Rechtsman (see the first figure). From a mathematical perspective, the model used to describe both systems consists of a discrete nonlinear Schrödinger (NLS) equation in two spatial dimensions, with periodic coefficients in the propagation variable. One-dimensional solitons in uniform waveguides, but without topology, were theoretically predicted in (9) and observed a decade later (10). They were subsequently observed in two-dimensional

uniform waveguides (11). These uniform systems are modeled by one- and two-dimensional discrete NLS equations with constant coefficients.

Researchers have theoretically predicted the existence of solitons (12) on the boundary edge of the helically varying waveguides used in the experiments of (7). The linear topology and the unidirectional propagation through defects appear to be naturally inherited by the nonlinear soliton modes (see the second figure). We anticipate future research that will continue to examine how the presence of topology affects the behavior of solitons.

An aspect of the work of Mukherjee and Rechtsman indicates the extent to which waveguide fabrication has progressed. Early fabricated optical structures created waveguides with uniform features in the late 1990s. This approach was extended to waveguides with helical structure in the longitudinal direction with femtosecond laser writing techniques in (7). The waveguides used in the present study are engineered so that during one period, each waveguide couples to its nearest neighbors sequentially and one at a time. All sorts of truly complex waveguides that could demonstrate noteworthy wave features can now be constructed, so this observation of topological solitons is likely to be one of many more. ■



REFERENCES AND NOTES

1. S. Mukherjee, M. C. Rechtsman, *Science* **368**, 856 (2020).
2. M. J. Ablowitz, *Nonlinear Dispersive Waves* (Cambridge Univ. Press, 2011).
3. N. J. Zabusky, M. D. Kruskal, *Phys. Rev. Lett.* **15**, 240 (1965).
4. D. J. Thouless, M. Kohmoto, M. P. Nightingale, M. den Nijs, *Phys. Rev. Lett.* **49**, 405 (1982).
5. F. D. M. Haldane, S. Raghu, *Phys. Rev. Lett.* **100**, 013904 (2008).
6. Z. Wang, Y. Chong, J. D. Joannopoulos, M. Soljačić, *Nature* **461**, 772 (2009).
7. M. C. Rechtsman *et al.*, *Nature* **496**, 196 (2013).
8. Y. Lumer, Y. Plotnik, M. C. Rechtsman, M. Segev, *Phys. Rev. Lett.* **111**, 243905 (2013).
9. D. N. Christodoulides, R. I. Joseph, *Opt. Lett.* **13**, 794 (1988).
10. E. Eisenberg, Y. Silberberg, R. Morandotti, A. R. Boyd, J. S. Aitchison, *Phys. Rev. Lett.* **81**, 3383 (1998).
11. J. W. Fleischer, M. Segev, N. K. Efremidis, D. N. Christodoulides, *Nature* **422**, 147 (2003).
12. M. J. Ablowitz, C. W. Curtis, Y.-P. Ma, *Phys. Rev. A* **90**, 023813 (2014).

ACKNOWLEDGMENTS

Supported by Air Force Office of Scientific Research grant FA9550-19-1-0084.

10.1126/science.abb5162

PLANT BIOLOGY

Breeding a fungal gene into wheat

An ancient cross-kingdom gene transfer enables wheat resistance to a fungal toxin

By Brande B. H. Wulff¹ and Jonathan D. G. Jones²

Every year, infection of wheat by the fungus *Fusarium graminearum* results in losses of ~28 million metric tons of wheat grain (1), valued at \$5.6 billion. The fungus reduces yields but also contaminates harvests with trichothecene toxins such as deoxynivalenol (DON; also called vomitoxin because of its effects on mammals) that render grain too poisonous to use. The disease is becoming more prevalent because of increasing cultivation of maize (also a host for the fungus) and reduced tillage (ploughing) agriculture, which promotes fungal survival on last season's plant debris. On page 844 of this issue, Wang *et al.* (2) reveal the molecular identity of the *Fusarium head blight 7* (*Fhb7*) gene, which encodes a glutathione *S*-transferase that detoxifies DON. This gene was acquired through a "natural" fungus-to-plant gene transfer in a wild wheat relative. This naturally occurring genetically modified (GM) wheat strain is therefore exempt from regulation and can be grown directly by farmers.

Annual yield losses due to *Fusarium* head blight are second only to leaf rust (1). Despite screening thousands of wheat lines, little resistance to *Fusarium* has been found. Wild grassy relatives of wheat, however, represent a rich source of genetic diversity, which has long been mined for resistance genes by interspecific crossing. The *Fhb7* gene was introduced into wheat from tall wheat grass (*Thinopyrum ponticum*) and provides major, semidominant resistance (2), unlike most *Fusarium* resistance in wheat, which is typically conferred by polygenic minor-effect genes that are difficult for breeders to track (3).

The identification of *Fhb7* by Wang *et al.* reveals an enzyme that detoxifies DON by

¹John Innes Centre, Norwich Research Park, Norwich, UK. ²The Sainsbury Laboratory, University of East Anglia, Norwich Research Park, Norwich, UK. Email: brande.wulff@jic.ac.uk; jonathan.jones@tsl.ac.uk

conjugating it to glutathione (see the figure). This explains the resistance conferred by *Fhb7* because DON is an important virulence factor required for *Fusarium* growth on infected tissue (4). One could now engineer *Fhb7* for DON detoxification to increase resistance to *Fusarium* species that cause head blight in other cereals (such as barley and rye) or crown rot in wheat and ear rot in maize.

The study of *Fusarium* head blight in wheat has been hindered by a disease resistance trait that is difficult to measure, a paucity of variation for resistance, and recent controversy concerning *Fhb1*, the first *Fusarium* head blight resistance gene to be cloned in wheat. Although one study identified *Fhb1* as a pore-forming toxin-like gene,

leaves, how DNA from *Epichloë* could enter the *Thinopyrum* germ line remains a mystery. The *Fhb7* gene was found to have 97% identity with its homolog in *Epichloë* but was otherwise absent from grass genomes, except within the *Thinopyrum* genus, suggesting that the gene transfer event arose after divergence of *Thinopyrum* from other grasses ~5 million years ago (2). Horizontal gene transfer events (the transfer of genetic material between species) are rare but have been recorded before, for example, between *Agrobacteria* and sweet potato (7) and between sorghum and parasitic *Striga* (8). In these cases, no beneficial function was associated with the transfer. Additional such horizontal gene transfers likely exist and might be revealed with bioinformatic

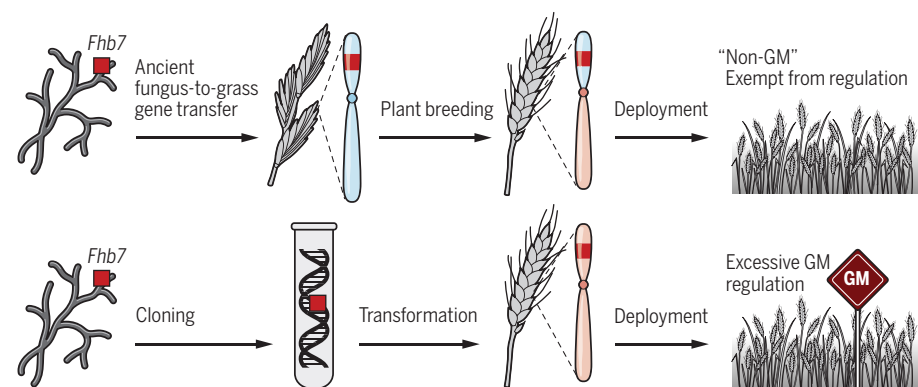
(aubergine), and sugar beet provide pest, disease, and herbicide resistance. In rice, many GM traits have now been approved (10). However, wheat—the world's most widely grown crop and a source of 20% of the calories and protein consumed by humankind—is a “GM orphan” (11).

Important opportunities are being missed by postponing GM wheat. Pests and diseases limit wheat production by ~20% globally (1). This number masks regional epidemics that can cause complete local crop failure, which is devastating for small-holder farmers in developing countries. It is now possible to rapidly discover and clone disease resistance genes from wild crop relatives (12) and engineer this resistance into domesticated varieties (13). Combinations (“stacks”) of multiple broad-spectrum resistance genes will likely provide durable disease resistance. With conventional breeding, such stacks would be almost impossible to create and maintain.

Can *Fhb7* be used as an example to sway public opinion on anti-GM arguments? If plant breeders can take advantage of a natural horizontal gene transfer such as *Fhb7* to reduce crop losses, why not a deliberate horizontal gene transfer for the same reason? The world is heading toward a projected population of 9.6 billion in 2050, and increases in crop yields are not keeping pace with growing demand. To meet this demand, and sustainably increase agricultural output, a concerted effort from breeders, agronomists, biotechnologists, and policymakers and effective public engagement from scientists about the “naturalness” of horizontal gene transfer is needed. ■

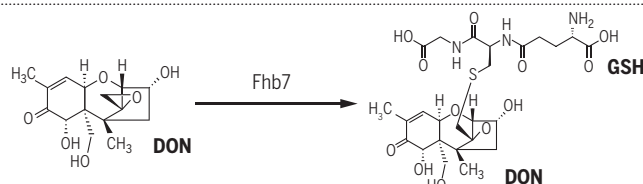
Two paths to *Fusarium*-resistant wheat

Fusarium head blight-resistant domesticated wheat has been produced by ancient horizontal transfer of the *Fusarium* head blight 7 (*Fhb7*) gene between *Epichloë*, a fungal endophyte, and wild wheatgrass. This gene could also be engineered into domesticated wheat but would be regulated as a genetically modified (GM) crop.



Mechanism of resistance

The *Fhb7* enzyme neutralizes deoxynivalenol (DON) by conjugating a glutathione (GSH) onto its toxic epoxide moiety.



two subsequent studies reported a histidine-rich calcium-binding protein but disagreed about the mode of action (5). Given the strong evidence presented by Wang *et al.*, including gain- and loss-of-function studies and a biochemical mechanistic dissection, hopefully *Fhb7* will evade such controversy. Optimal control of wheat head blight may require breeders to combine *Fhb7* with *Fhb1*, but this remains to be rigorously tested.

The most extraordinary aspect of *Fhb7* concerns its origin in *Epichloë*, a widely distributed ascomycete fungal genus that colonizes leaves of many grasses. Some species make alkaloid neurotoxins that render ryegrass poisonous to sheep in New Zealand (6). Because *Epichloë* primarily colonizes

searches. Moreover, why *Epichloë* evolved a DON detoxification gene is unknown; perhaps it detoxifies one of its own toxins or helps *Epichloë* compete with *Fusarium* for grass colonization.

What does the natural transfer of *Fhb7* into a grass mean for the discussion on GM crops? This natural GM product may be as good as or better than any that could have been created in the laboratory (see the figure), although conceivably, *Fhb7* could be even more effective if highly expressed from other promoters (2). Despite concerns from some, GM crop cultivation is increasing. Of the world's arable land, 10% is used for GM soy, maize, cotton, and canola (9), which along with GM potato, papaya, eggplant


REFERENCES AND NOTES

1. S. Savary *et al.*, *Nat. Ecol. Evol.* **3**, 430 (2019).
2. H. Wang *et al.*, *Science* **368**, eaba5435 (2020).
3. M. Buerstmayr *et al.*, *Plant Breed.* (2019). 10.1111/pbr.12797
4. G.-H. Bai *et al.*, *Mycopathologia* **153**, 91 (2002).
5. E. S. Lagudah, S. G. Krattiger, *Nat. Genet.* **51**, 1070 (2019).
6. R. T. Gallagher, E. P. White, P. H. Mortimer, N. Z. Vet. J. **29**, 189 (1981).
7. T. Kyndt *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 5844 (2015).
8. S. Yoshida, S. Maruyama, H. Nozaki, K. Shirasu, *Science* **328**, 1128 (2010).
9. <https://royalsocietypublishing.org/~/media/policy/projects/gm-plants/gm-plant-q-and-a.pdf>
10. www.isaaa.org/gmapprovaldatabase
11. B. B. H. Wulff, K. S. Dhugga, *Science* **361**, 451 (2018).
12. S. Arora *et al.*, *Nat. Biotechnol.* **37**, 139 (2019).
13. <http://2blades.org/2019/03/12/wheat-lines-from-2blades-csiro-and-umn-exhibit-exceptional-stem-rust-resistance-in-the-field>

ACKNOWLEDGMENTS

The authors are supported by the Biotechnology and Biological Sciences Research Council (B.B.H.W. and J.D.G.J.), the 2Blades Foundation (B.B.H.W. and J.D.G.J.), and the Gatsby Foundation (J.D.G.J.). We thank G. Brar for helpful discussions.

10.1126/science.abb9991



A bumble bee (*Bombus terrestris*) worker is poised to damage a leaf to spur flowering, which suggests a phenological synchrony between plants and pollinators.

ECOLOGY

The secret lives of bees as horticulturists?

Pollen-starved bumble bees may manipulate plants to fast-forward flowering

By **Lars Chittka**

The symbiotic relationship between insect pollinators and flowers is as tight as it is fragile. Plants provide nutrition for flower visitors in exchange for pollination services. However, timing is critical in this exchange: If pollinators are out of sync with the blooming times of their favorite flowers, then the plants might display their beautiful sex organs in vain, and pollinators and their offspring might starve (1, 2). Now, on page 881 of this issue, Pashalidou *et al.* report that bumble bees are not passive players in this relationship.

School of Biological and Chemical Sciences,
Queen Mary University of London, London, UK.
Email: l.chittka@qmul.ac.uk

Rather, the bees have developed a way to hasten flowering specifically under conditions in which bees are food-deprived early in the season (3).

Climate change threatens the long-established synchronization of seasonal pollinator activity and flowering time. Temperature strongly affects the emergence of pollinating insects after hibernation (4). By contrast, flowering relies heavily on the time of exposure to light (the photoperiod), which is not subject to climate change (5). Thus, pollinators might find themselves critically short of nutrition early in the season.

In contrast to hypothesis-driven research, scientific discovery often springs from careful observation of natural phenomena. Pashalidou *et al.* noticed a previously unreported natural behavior: Using their mouth

parts, bumble bees deliberately damaged leaves of a variety of plant species (see the figure). The authors suspected that this behavior might be related to a shortage of pollen, the bees' sole source of protein. Therefore, the researchers compared the leaf-damaging behavior exhibited by experimental pollen-starved colonies of bumble bees with that of worker bees from well-fed nests. The results were consistent across years and experimental situations (laboratory settings as well as free-flying colonies): Pollen-starved workers made considerable efforts to puncture holes in the leaves of flowerless plants, whereas workers from well-fed colonies rarely did so.

But why? Pashalidou *et al.* then discovered a dramatic effect of the leaf-damaging behavior on flowering phenology. When exposed to leaf-biting bees, the black mustard plant

PHOTO: HANNER PULIDO DE MORAES AND MESCHER LABORATORIES

(*Brassica nigra*) flowered 2 weeks earlier and tomato plants (*Solanum lycopersicum*) came into flower a month sooner than would normally be expected. Thus, bumble bees appear to perform a low-cost, but highly efficient, trick to accelerate flowering in plants around their nest under conditions when flower pollen resources are most urgently needed for colony growth.

Many intriguing questions surround the evolution of leaf-biting behavior as well as its adaptive importance. How might this behavior have arisen? One possibility is that individual bees figured out that leaf-biting results in future rewards, and that these bees remember the very plants they have damaged and return to them weeks later to reap the benefits of their efforts. This is perhaps not wholly implausible, given that bumble bees have developed other impressively innovative solutions to access rewards (6), and their spatial memories can last a lifetime (7, 8). However, it is unlikely that bees can learn that a link exists between an action and a reward that occurs a month later. Also, worker bees in the wild rarely survive longer than 1 month (7).

An alternative explanation for how leaf-biting first arose is that individual bees receive an immediate benefit in addition to the more long-term one for colony fitness. For example, bees might extract a substitute protein source from leaf-biting, such as plant sap. However, Pashalidou *et al.* rejected this possibility because most leaf-damaging interactions seemed too brief for bees to imbibe plant juices in appreciable quantities.

Perhaps pollen-starved bees just bite plant parts indiscriminately in the hope that these might conceal some pollen. This too is unlikely, because even entirely inexperienced bees can tell flowers from vegetative parts (9). Bumble bees sometimes extract nectar from hard-to-access flowers by puncturing floral structures, a technique called nectar robbing. Inexperienced workers attempt this at various flower parts, until they figure out the reward location (10).

One might also wonder why bees would bite holes in vegetative parts of plants that do not even have flowers, instead of searching for plants that do. Unselective perforation of leaves in a bee colony's flight range will not confer much profit. For example, spreading the perforation treatment too far from the native nest might be more likely to benefit competing bees with nearby nests. In addition, many plants, such as mosses or ferns, will never provide any useful pollen for bees, nor will it be beneficial to bite the leaves of plants that are past the blooming stage. Perhaps bumble bees can use flower buds as cues that flowering is on the horizon. Thus, the bees would know that these plants are

worth their effort to further speed up flowering (11). Future studies should develop a plausible evolutionary scenario for how the first mutant bees that began leaf-biting might have conferred a sufficient selective colony-fitness advantage for this trait to spread through a population.

Turning to plants, there are many equally fascinating questions relating to why an adequate response to bee-driven leaf damage would be to accelerate flower development. One possibility is that such damage is interpreted by the plants as an ongoing herbivore attack; annual plants, therefore, might force an earlier flowering period before the plant's untimely demise. Plants are known to speed up their flowering as a response to various stressors, but there are no known examples of such a response to herbivory (4). An adaptive explanation might be that plants "want" to respond to bees that are signaling a dearth of food, because this also means there might be an excess of pollination services available. However, there will also be an opposing selective pressure to synchronize flowering with potential mates within plant species (12), which would be a disadvantage to individual plants that move their flowering forward.

A further reason why plants might fast-track flowering is that they are simply manipulated into doing so against their own advantage, but to the benefit of bumble bees. Mechanical damage made with metal forceps and razors does not have the same effect on flowering times as does perforation by bees. Thus, it remains possible that bees inject chemicals into the plants to promote flowering. If so, scientists might realize a horticulturist's dream by deciphering the molecular pathways through which flowering can be accelerated by a full month. An encouraging interpretation of the new findings is that behavioral adaptations of flower visitors can provide pollination systems with more plasticity and resilience to cope with climate change than hitherto suspected. ■

REFERENCES AND NOTES

1. J. Memmott, P. G. Craze, N. M. Waser, M. V. Price, *Ecol. Lett.* **10**, 710 (2007).
2. J. D. Thomson, *Philos. Trans. R. Soc B* **365**, 3187 (2010).
3. F. G. Pashalidou, H. Lambert, T. Peybernes, M. C. Mescher, C. M. De Moraes, *Science* **368**, 881 (2020).
4. K. Takeno, *J. Exp. Bot.* **67**, 4925 (2016).
5. R. J. Stelzer, L. Chittka, M. Carlton, T. C. Ings, *PLOS ONE* **5**, e9559 (2010).
6. S. Alem *et al.*, *PLOS Biol.* **14**, e1002564 (2016).
7. J. L. Woodgate, J. C. Makinson, K. S. Lim, A. M. Reynolds, L. Chittka, *PLOS ONE* **11**, e0160333 (2016).
8. L. Chittka, *J. Exp. Biol.* **201**, 515 (1998).
9. K. Lunau, *Ethology* **88**, 203 (1991).
10. E. Leadbeater, L. Chittka, *Proc. Biol. Sci.* **275**, 1669 (2008).
11. J. A. Rosenheim, *Behav. Ecol. Sociobiol.* **21**, 401 (1987).
12. J. Forrest, J. D. Thomson, *Am. J. Bot.* **97**, 38 (2010).

ACKNOWLEDGMENTS

The author thanks J. Thomson for discussions.

10.1126/science.abc2451

METROLOGY

Synchronized to an optical atomic clock

Microwave generation using optical frequency comb technology hits new milestones

By E. Anne Curtis

Metrology, the science of measurement, is at the heart of all scientific endeavors. Of all parameters, frequency is the most accurately measurable quantity in the scientific portfolio. Tabletop experiments can generate highly accurate and stable frequencies that are being used to challenge the tenets of fundamental physics (1) as well as for specific applications such as the synchronization of large-scale radio telescope arrays (2). Optical atomic frequency standards have the intrinsic capacity to attain higher levels of stability and accuracy than microwave-based standards. Microwave technology, used in every sector of society, would benefit greatly from similar performance. On page 889 of this issue, Nakamura *et al.* introduce an experimental system with the ability to transfer the precise phase and accuracy of optical clock signals into the electronic domain, while demonstrating a fractional frequency instability of one part in 10^{18} (3). This result brings the superior performance of an optical frequency standard into the microwave regime.

The innovation of the femtosecond optical frequency comb (OFC) was a major breakthrough in the pursuit of improved frequency standards based on optical transitions (4). Although optical signals oscillate much too quickly for their frequencies to be counted electronically, an OFC can phase-coherently link optical frequencies to much slower microwave frequencies. The OFC is best described as a pulsed-laser device whose output is a series of very short-lived light pulses, produced at a repetition rate in the microwave regime. The output can also be observed in the frequency domain, where it looks like a comb of evenly spaced fre-

Time & Frequency Department, Optical Frequency Metrology, Atomic Clocks & Sensors, Quantum Metrology Institute, National Physical Laboratory, Teddington, Middlesex TW11 0LW, UK. Email: anne.curtis@npl.co.uk

quencies or tones spanning the optical regime. The repetition rate can be referenced to the microwave-based International System of Units (SI) definition of the second. Because of a direct relation between the repetition rate and the comb tone spacing, the absolute frequency of each tone can be determined. Measuring the optical beat between a single comb tone and an unknown optical frequency provides a method for directly linking optical frequencies to the microwave reference. The inverse should also be true, as the stability and accuracy provided by an optical frequency standard can be used as a reference for the OFC. This means locking the comb in step with the optical oscillations. The comb is then expected to produce an equally stable output in the microwave regime (see the figure), and the measurement of this seemed the obvious next step (5). The devil was in the details, and it took nearly 20 years to come full circle to obtain exceptional accuracy and long-term stability in microwave generation.

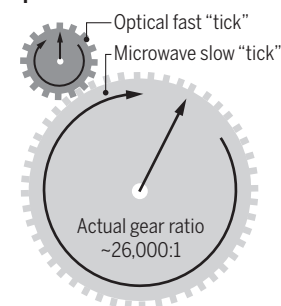
Stable and accurate microwave signals are used in a variety of arenas including communication, navigation, radar, radio astronomy, and fundamental physics research. Currently, the most accurate and stable frequency comparison and dissemination is achieved with ultrastable lasers transmitted by optical fiber networks. These systems are being used to lay the groundwork for a redefinition of the SI second by enabling direct optical-to-optical comparisons of frequency standards separated by large geographical distances. Furthermore, frequency transfer networks are used in probes of fundamental physics (6) and detection of submarine earthquakes by means of deep-sea fiber optic cables (7), among other applications.

Maintaining the stability and accuracy of microwaves over long distances, whether for synchronization, communication, or navigation, requires overcoming some obstacles in transmission and signal processing. Atmospheric and delay compensation has improved—for example, with two-way satellite time and frequency transfer for satellite- and ground-based systems. Space-based applications using ultrastable microwaves avoid the issues of atmospheric disturbance altogether. By using the ultralong baselines only accessible to space-based systems, radio telescope arrays in space synchronized by ultrastable

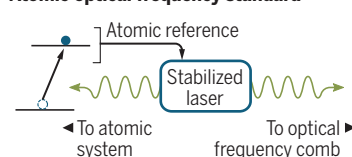
Making microwaves

An optical frequency standard uses a signal generated from an atomic excitation to steer a stabilized laser output. Combining this with the output of an optical frequency comb produces an interference beat. Stabilizing the beat frequency forces the repetition rate of the comb to lock in step with the frequency reference, only with a microwave output.

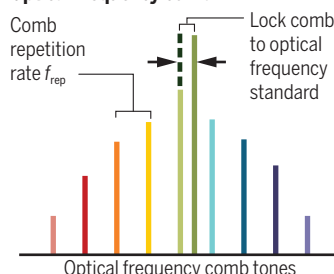
Optical-to-microwave conversion



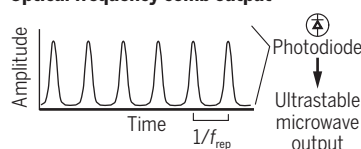
Atomic optical frequency standard



Optical frequency comb



Optical frequency comb output



microwaves have the potential to greatly increase the resolution of the signals (8).

Unlike for satellite-based transmission, Doppler radar systems can immediately take advantage of stability and accuracy improvements in microwave reference signals. These radar systems simply compare an outgoing signal and a received echo, using a single common clock signal and path. Upgraded microwave stability directly increases the attainable accuracy, unblurring the images of distant and moving targets (9). This improved accuracy will help in efforts to upgrade navigation and tracking systems across multiple sectors.

The directed development of new technologies as a result of scientific need runs through several recent innovations important for improved accuracy and stability. An essential component for the optical-to-electrical conversion through OFC is the optical detector that turns the regularly spaced output pulses of light into a signal of ultrastable microwaves. Extensive and focused research and development was required to produce high-speed photodiodes with high linearity even when processing high peak intensities in the pulse train. Exceptionally stable microwave generation and measurement required exploiting the growing field of software-defined radio (10), which combines multiple systems of microwave hardware devices into one reconfigurable software-driven device. This approach is gaining widespread use for frequency production and manipulation. This trend highlights a move from analog to

digital frequency synthesis and for integration of signal generation and processing. Finally, innovations in the OFC technology itself were key to improved microwave generation and can now be found in commercially available devices.

Miniaturization and integration of several technologies will be necessary to enable widespread stable microwave and optical frequency production and dissemination. The next step likely involves the continued development of portable atomic clocks and optical frequency “microcombs” (microOFCs) (11). Far-ranging operation could involve optical clock-level time and frequency provided by fully integrated systems of miniature optical clocks, microOFCs, and noise-reducing detectors and microwave electronics. Because optical clocks have achieved unprecedented levels of accuracy (12) and stability (13), linking the frequencies provided by these optical standards with distantly located devices would allow direct calibration of microwave clocks to the future optical SI second. This combination would markedly improve synchronization capabilities between multiple locations and devices, and over large distances.

Like the revolution that OFC technology produced in the field of optical frequency metrology, Nakamura *et al.*'s ability to produce microwaves with the stability and accuracy afforded by an optical clock signal is a paradigm shift in the field of microwave metrology. The impact will extend to applications in fundamental physics, communication, navigation, and microwave engineering. Growing access to this new frontier of ultrastable microwave sources will only push these sectors and others to new innovative heights. ■

REFERENCES AND NOTES

1. A. D. Ludlow, M. M. Boyd, J. Ye, E. Peik, P. O. Schmidt, *Rev. Mod. Phys.* **87**, 637 (2015).
2. C. Clivati *et al.*, *Sci. Rep.* **7**, 40992 (2017).
3. T. Nakamura *et al.*, *Science* **368**, 889 (2020).
4. T. Fortier, E. Baumann, *Commun. Phys.* **2**, 153 (2019).
5. S. A. Diddams *et al.*, *Science* **293**, 825 (2001).
6. P. Delva *et al.*, *Phys. Rev. Lett.* **118**, 221102 (2017).
7. G. Marra *et al.*, *Science* **361**, 486 (2018).
8. L. I. Gurvits, *Adv. Space Res.* **65**, 868 (2020).
9. P. Ghelfi *et al.*, *Nature* **507**, 341 (2014).
10. J. A. Sherman, R. Jördens, *Rev. Sci. Instrum.* **87**, 054711 (2016).
11. P. Del'Haye *et al.*, *Nat. Photonics* **10**, 516 (2016).
12. S. M. Brewer *et al.*, *Phys. Rev. Lett.* **123**, 033201 (2019).
13. E. Oelker *et al.*, *Nat. Photonics* **13**, 714 (2019).

10.1126/science.abb8629

Restoring vision to the blind

Ideas abound to restore vision to people blinded by retinal disease

By John E. Dowling

Surveys consistently report that people fear total blindness more than any other disability, and currently the major cause of untreatable blindness is retinal disease. The retina, a part of the brain that extends into the eye during development, initiates vision by first detecting light with the rod and cone photoreceptors. Four classes of retinal neurons then begin the analysis of visual images. Defects in the optical media that transmit and focus light rays onto the retina (lens and cornea) can usually be dealt with surgically, although such treatments are not available in some parts of the world, resulting in as many as 20 to 30 million legally blind individuals worldwide. Untreatable retinal disease potentially causes legal or total blindness in more than 11 million people in the United States alone, but progress in treatments raises the possibility of restoring vision in several types of retinal blindness (1).

Retinal neurons comprise bipolar and horizontal cells, which are second-order neurons that receive signals from the photoreceptors in the outer retina. Third-order amacrine and retinal ganglion cells are activated in the inner retina by bipolar cells. Axons from the ganglion cells form the optic nerve and carry the visual message to the rest of the brain (see the figure). The cells most susceptible to blinding retinal disease are the photoreceptors and ganglion cells. Whereas progress has been made in combating blindness caused by photoreceptor degeneration, little can be done currently to address ganglion cell loss, such as occurs in glaucoma.

The approach that has been most successful in restoring photoreceptor loss that results in complete blindness is the use of retinal prosthetic devices, with two now approved for clinical use (2). These devices electrically stimulate either bipolar or ganglion cells. They require goggles that have a camera that converts visual stimuli into electrical stimuli that activate the device, which in turn stimulates the retinal cells. Several hundred of these devices have

been implanted in blind or virtually blind individuals, 70 to 80% of whom report improvement in quality of life. For those who are completely blind, the ability to experience again at least some visual function is viewed as a miracle.

There are substantial limitations to the devices, however. The best visual acuity attained so far is poor (20/500) and visual field size is limited, but many improvements, mainly technical, are being developed and tested, including the potential use of electronic low-vision devices to increase visual field size and acuity (3). Retinal prostheses are not useful for patients who are blind because of loss of ganglion cells and/or the optic nerve, but prostheses that bypass the retina and stimulate more central

“...progress in treatments raises the possibility of restoring vision in several types of retinal blindness.”

visual structures, including the lateral geniculate nucleus (the intermediary between retina and cortex) and visual cortex, are being developed and tested in humans (4). There remain considerable technical issues, but preliminary data indicate that such devices are feasible.

A second approach to treat photoreceptor degeneration and potential blindness, now in the clinic, is gene therapy (5). This involves injecting a viral construct into the eye that contains a normal gene to replace an abnormal one. Success so far has been limited to the treatment of Leber congenital amaurosis (LCA) type 2, a rare form of retinitis pigmentosa in which the gene whose product is required to form the correct isomer of vitamin A aldehyde, the chromophore of the visual pigments, is mutated. Little of the correct isomer is made in LCA patients, resulting in substantial loss of photoreceptor light sensitivity. This is reversed when viral constructs encoding the normal gene are injected deep into the eye between the photoreceptors and pigment epithelium.

Two factors make this approach feasible in LCA: The genetic defect is monogenic, and many of the photoreceptors in the patients remain alive, although compromised. Thus, how broadly feasible gene therapy

will be for treating the enormous range of inherited retinal diseases now known to exist (~300) remains to be seen. But at least a dozen other gene therapy trials on monogenic inherited eye diseases similar to LCA are under way (6). Other methods to manipulate genes are now available, including CRISPR-mediated editing of retinal genes. So far, the experiments have been mainly on isolated cells or retinas, but these powerful techniques are likely to have eventual clinical applications.

A variation on the use of gene therapy techniques is optogenetics, in which light-sensitive molecules are introduced into non-photosensitive retinal cells. This approach holds much promise for restoring vision to totally blind individuals whose photoreceptors have been lost. Using viruses to insert genes encoding light-sensitive molecules into bipolar and ganglion cells, as well as surviving photoreceptor cells that are no longer photosensitive, has been accomplished in animals and shown to restore some

vision (7). Again, technical issues remain: The cells made light-sensitive require bright light stimuli, and the light-sensitive cells do not adapt. That is, whereas photoreceptors normally allow vision over as much as 10 log units of light intensity, the cells made light-sensitive respond only to a range of 2 to 3 log units. Various methods to overcome these limitations are now being developed, and at least one clinical trial is under way. Experiments to make cortical neurons sensitive to light or other stimuli that better penetrate the skull—magnetic fields or ultrasound, for example—are also being developed and tested in animals.

Other promising approaches to restore vision are being explored. In cold-blooded vertebrates, retinal cells (in fish) and even the entire retina (in amphibians) can regenerate endogenously after damage. Regeneration of retinal cells in zebrafish is now quite well understood (8). The regenerated neurons come from the major glial cell in the retina, the Müller cell. After retinal damage, Müller cells reenter the cell cycle and divide asymmetrically to self-renew and produce a progenitor cell that proliferates to produce a pool of cells capable of differentiating into new retinal cells that repair the retina.

A number of transcription factors and

Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA. Email: dowling@mcb.harvard.edu

other factors identified as being involved in retinal regeneration in zebrafish have been shown to stimulate some Müller cell proliferation and neuronal regeneration in mice. Regenerated bipolar and amacrine cells, as well as rod photoreceptors, have so far been identified in mouse retinas, and these cells are responsive to light stimuli (9, 10). Further, cells postsynaptic to the regenerated neurons are activated by light stimuli, indicating that the regenerated neurons have been incorporated into the retinal neural circuitry. So far, the regenerative capacity of mammalian Müller cells is limited, but directed differentiation of specific types of neurons with a mix of factors appears to be a possibility. Regrowth of ganglion cell

suggests that only a small proportion of the donor cells integrate, but progress in overcoming this setback is being made.

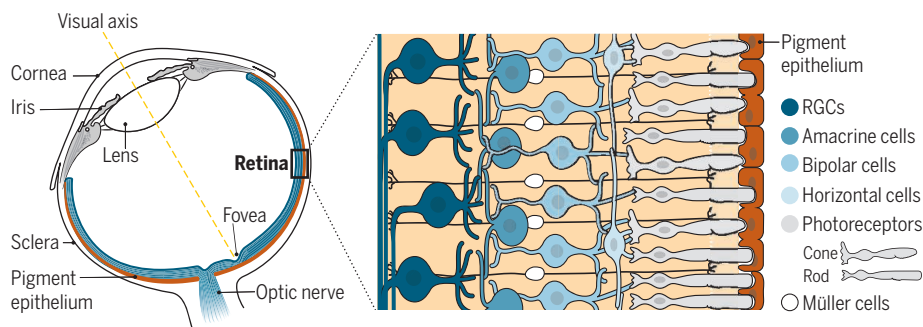
More success has been reported with stem cells induced to become pigment epithelial (PE) cells, which provide essential support for photoreceptors. A number of blinding retinal diseases relate to the degeneration of the PE cells, and replacement using such cells—in a suspension or on a scaffold—is being actively pursued. PE cells do not need to integrate synaptically with retinal cells; they simply need to contact the photoreceptor cells. This is achieved when PE cells are placed between the retina and the back of the eye. Early clinical trials suggest that the transplants are safe, but reti-

Two of the most common retinal diseases in developed countries—age-related macular degeneration (AMD), the leading cause of legal blindness (visual acuity of less than 20/200), and glaucoma, the leading cause of total blindness—are not monogenic diseases, and so genetic treatments for them are not obvious. Attempts to understand the etiology of these diseases are under way, but currently their underlying causes are still unclear. A difficulty presented by AMD is that no animal model is readily available, because it is a disease of the fovea, which mediates high-acuity vision. Except for primates, other mammals do not possess this small critical retinal area. Whereas large primates are not feasible for extensive cellular or molecular studies, small primates such as marmosets that have a fovea are potential models but have not been used much to date.

Other approaches for restoring vision have been suggested and have even yielded some progress. From both normal humans and those with an inherited retinal disease, skin biopsy cells can be induced to form tiny retinal eyecups called organoids (15). Containing all retinal cell types, these structures could be a source of retinal cells for studying retinal disease development and possible therapies, as well as for cell transplantation. A fovea has not been observed in any organoid so far, but this is not beyond the realm of possibility. Another suggested approach is to surgically transplant whole eyes into blind individuals. This appears feasible, but whether there is sufficient optic nerve regrowth remains an open question. ■

The eye and retina

The retina lines the back of the eye and consists of rod and cone photoreceptors, as well as four types of neuron: second-order bipolar and horizontal cells and third-order retinal ganglion cells (RGCs) and amacrine cells. Müller glial cells fill the spaces between the neurons. The pigment epithelium, critical for photoreceptor function, underlies the retina. Photoreceptors and RGCs are most susceptible to blinding retinal disease. Progress in combating photoreceptor degeneration has been made, but there are few strategies to address RGC loss.



axons after the optic nerve is disrupted is also under active investigation, and although the number of axons regrowing is low (~10%), those that do regrow establish synaptic connections with their correct targets (11). Therefore, endogenous regeneration is still far from clinical testing, but substantial progress has occurred.

A long-studied area of research is transplantation of retinal cells, particularly photoreceptors, into diseased retinas. In experiments with mice, transplanted post-mitotic rod photoreceptor precursor cells derived from embryonic retinas or from stem cells appeared to integrate into diseased retinas in reasonable numbers and to be functional. A surprising and unexpected complication in the interpretation of these experiments was recently discovered. Rather than integrating into diseased retinas, the donor cells appear to pass material (RNA or protein) into remaining host photoreceptor cells, rejuvenating them, and these appear to be most of the functional cells (12). The current evidence

nal detachment, a serious complication, can occur and efficacy has yet to be shown (13).

The finding that donor photoreceptor cells can help diseased host retinal cells to recover function suggests that certain substances can provide neuroprotection. Indeed, a substantial number of such neuroprotective molecules have been shown to affect retinal disease progression, especially degeneration of photoreceptor cells. No one factor has been shown to be effective generally, but two have received much attention. One, ciliary neurotrophic factor (CNTF), promotes photoreceptor survival in light-induced photoreceptor degeneration and in several other models of retinal degeneration (14). Some evidence suggests that CNTF acts primarily on Müller cells, but how it works, and on what cells, is still unclear. The other factor, rod-derived cone viability (RCDV) factor, has received less research attention, but with recent industrial support, it is now being advanced to the clinic. Current evidence indicates that RCDV factor protects cones after rod degeneration.

REFERENCES AND NOTES

1. J. E. Dowling, K. M. Wright, Eds., special issue on Restoring Vision to the Blind. *Trans. Vis. Sci. Tech.* **3**(7) (2014).
2. J. O. Mills, A. Jalil, P. E. Stanga, *Eye* **31**, 1383 (2017).
3. A. D. Deemer *et al.*, *Optom. Vis. Sci.* **95**, 694 (2018).
4. E. Fernández, R. A. Normann, in *Artificial Vision*, V. P. Gabel, Ed. (Springer, 2015), pp. 191–201.
5. S. Russell *et al.*, *Lancet* **390**, 849 (2017).
6. A. V. Garafalo *et al.*, *Prog. Ret. Eye Res.* **10**, 1016/jpretyeres.2019.100827 (2019).
7. J.-A. Sahel, J. Bennett, B. Roska, *Sci. Transl. Med.* **11**, eaax2324 (2019).
8. J. F. Martin, R. A. Poché, *Development* **146**, dev182642 (2019).
9. N. L. Jorstad *et al.*, *Nature* **548**, 103 (2017).
10. F. Bei *et al.*, *Cell* **164**, 219 (2016).
11. S. G. Varadarajan, A. D. Huberman, *Curr. Opin. Neurobiol.* **53**, 198 (2018).
12. R. A. Pearson *et al.*, *Nat. Commun.* **7**, 13029 (2016).
13. M. S. Mehat *et al.*, *Ophthalmology* **125**, 1765 (2018).
14. E.-Y. Chew *et al.*, *Ophthalmology* **126**, 540 (2019).
15. E. E. Capowski *et al.*, *Development* **146**, dev171686 (2019).

ACKNOWLEDGMENTS

I am grateful to the participants of the Lasker/IRRF study on Restoring Vision to the Blind (1) who provided recent findings in the field.

10.1126/science.aba2623

Rapid repurposing of drugs for COVID-19

The emergence of a new coronaviral respiratory disease calls for repurposing existing drugs

By **R. Kiplin Guy¹, Robert S. DiPaola², Frank Romanelli¹, Rebecca E. Dutch²**

In late fall 2019, a novel acute respiratory disease, called coronavirus disease 2019 (COVID-19) emerged in Wuhan, China. COVID-19 is caused by severe acute respiratory syndrome-coronavirus 2 (SARS-CoV-2) (1, 2). COVID-19 has been declared a pandemic by the World Health Organization and continues to spread across the globe. Most patients recover within 1 to 3 weeks. However, a small proportion (~5%) develop severe illness that can progress to acute respiratory distress syndrome (ARDS), which can lead to death. Currently, only supportive care is available; patients would greatly benefit from the availability of direct therapeutic approaches. One approach to identifying therapeutics is to repurpose approved drugs developed for other uses, which takes advantage of existing detailed information on human pharmacology and toxicology to enable rapid clinical trials and regulatory review.

The coronaviruses are single-stranded RNA viruses that infect vertebrates and move between different host species (3). With the emergence of SARS-CoV-2, there are now seven coronaviruses that are known to infect humans. Four of them (HCoV-229E, HCoV-OC43, HCoV-NL63, and HCoV-HKU1) are responsible for ~30% of cases of the common cold in humans. Two of them caused recent epidemics that had considerable associated mortality: SARS-CoV-1, which emerged in 2002–2003 and causes ~10% mortality, and Middle East respiratory syndrome coronavirus (MERS-CoV), which emerged in 2012, is still active, and causes ~35% mortality. Both epidemics affected a relatively small number of patients compared with COVID-19, which is more transmissible for several reasons, including asymptomatic carriers, long latency period, and high infectivity. Before COVID-19, only SARS-CoV-1 and MERS-CoV caused severe disease. Therefore, coronaviral drug discovery has been a small effort relative to that for other viral diseases such as influenza. Given the rapid spread of COVID-19 and its

relatively high mortality, filling the gap for coronavirus-specific drugs is urgent.

The coronavirus life cycle (see the figure) involves a number of potentially targetable steps, including endocytic entry into host cells [involving angiotensin-converting enzyme 2 (ACE2) and transmembrane protease serine 2 (TMPRSS2)], RNA replication and transcription [involving helicase and RNA-dependent RNA polymerase (RdRp)], translation and proteolytic processing of viral proteins (involving chymotrypsin-like and papain-like proteases), virion assembly, and release of new viruses through the exocytic systems (4). In addition to virally encoded targets, numerous host targets are essential for viral replication and disease progression (3).

The cellular receptor for SARS-CoV-2 is ACE2 (5). Recombinant human ACE2 (rhACE2, or APN01) is currently under development as a treatment for acute lung injury and pulmonary arterial hypertension and has proven well tolerated in a phase 1 trial in healthy volunteers. rhACE2 has been shown to significantly reduce viral entry into human cell-derived organoids (6), presumably by acting as a decoy for virus binding. This has lent support to the clinical trials that are investigating blockade of viral entry with APN01 for COVID-19 patients. Successful viral entry requires proteolytic processing of the viral coat spike glycoprotein (S), which can be carried out by TMPRSS2 (7). The TMPRSS2 inhibitor camostat (7) is approved in Japan for the treatment of chronic pancreatitis and post-operative gastric reflux and is generally well tolerated, although rare serious side effects have been reported. Both camostat and the related agent nafamostat (8) block SARS-CoV-2 replication in TMPRSS2-expressing human cells. Camostat has been shown to block infection with SARS-CoV-2 in a mouse model. Therefore, there is a strong rationale to support clinical trials with these drugs for COVID-19, which have already been initiated in the Netherlands and Germany.

Coronaviruses use the endolysosomal pathway to enter the cell before uncoating. Chloroquine (CQ) and hydroxychloroquine (HCQ) are antimalarial drugs that affect endosomal function and block autophagosome-lysosome fusion (9). Both drugs have been shown to inhibit SARS-CoV-2 replication in cellular models (8, 10). Azithromycin

(AZ), a widely used broad-spectrum antibiotic, also blocks autophagosome clearance in human cells (11) and replication of the Zika virus and influenza virus in human cells in vitro (12). Preliminary results from a small randomized trial of HCQ in COVID-19 patients report a reduction in time to clinical resolution (13). A small open-label trial has demonstrated increased reduction in viral load for COVID-19 patients receiving the combination of HCQ and AZ relative to HCQ alone, although this study has been heavily criticized because of post hoc removal of several subjects from the study analysis (14). These hypothesis-generating studies have justified emergency approval of their use for COVID-19 in the United States, where they are both being widely used.

However, both HCQ and AZ have potential cardiac toxicity (QT prolongation, which can lead to fatal arrhythmia), and HCQ additionally has the potential for negative effects on the eye. Understanding risk-benefit ratios is paramount if these drugs are to become a standard of care for COVID-19. Several post hoc analyses carried out in the United States and Europe suggest modest benefit, at best, from HCQ monotherapy for COVID-19 patients; one large post hoc analysis among U.S. veterans suggests that there is harm to patients from HCQ. Given the mechanistic rationale but lack of well-designed clinical studies and potential for drug-induced toxicity, there is a key need for controlled, randomized trials to test the efficacy and safety of these drugs for COVID-19 patients.

After uncoating, the viral genomic RNA is used for cap-dependent translation to produce two polypeptides, which are then autoproteolytically processed to produce several viral proteins, including RdRp and two proteases. Although the proteases might seem attractive targets given the number of viral protease inhibitors previously developed for HIV and other viruses, they are only distantly related to other viral proteases. The combination of the HIV protease inhibitors lopinavir and ritonavir (15) proved clinically ineffective for COVID-19 patients, as had previously been the case for the same combination in SARS-CoV-1 disease. Therefore, further repurposing with this class of drugs is poorly justified—although there are other protease inhibitors in early-stage drug discovery that are directed to the coronavirus proteases.

¹College of Pharmacy, University of Kentucky, Lexington, KY, USA. ²College of Medicine, University of Kentucky, Lexington, KY, USA. Email: kip.guy@uky.edu

Production of the replication complex proteins, including the helicase and RdRp, allows for genomic replication of the virus and for production of subgenomic RNAs, which are also translated to produce structural and coat proteins. The helicase is theoretically an attractive target, but it is divergent from other viral helicases, and there is no evidence that the herpes simplex virus helicase inhibitors amenamevir or pretelivir are effective against coronaviruses.

RdRp carries out both replication and transcription of the viral RNA, making it a clear target for blocking the viral life cycle. Because RdRp is a critical protein for many viruses, a number of broad-spectrum RdRp inhibitors are either approved or in clinical trials, including remdesivir and favipiravir. Remdesivir was initially developed to treat the flaviviruses that cause Ebola and Marburg diseases and has proven safe in trials during the past two Ebola epidemics. However, it is less effective for Ebola than antibody-based treatments that prevent the virus from entering human cells. Remdesivir was subsequently shown to be active against both SARS-CoV-1 and MERS-CoV in animal models. Favipiravir was developed for influenza and approved in Japan in 2014, specifically for new pandemic influenza outbreaks. Both remdesivir and favipiravir are active against SARS-CoV-2 in human cells *in vitro* (7). Remdesivir has been rapidly advanced into several clinical trials for COVID-19, and early informal data being released from those trials suggest that remdesivir is effective, but such datasets need to be used cautiously for generalizing the understanding of either safety or efficacy. Further randomized, controlled trials with RdRp inhibitors are justified and needed.

The best justified drugs for repurposing to treat COVID-19 patients are the host-factor-targeted drugs HCQ, AZ, and camostat and nafamostat and the viral RdRp-targeted drugs remdesivir and favipiravir. A number of other drugs are also being considered, although with less supporting evidence (see supplementary materi-

als). Additionally, phenotypic screening approaches are being developed on the basis of either viral entry or replication that could be used to survey approved drugs and drug candidates much more widely. Both of these approaches may widen the available classes of drugs for consideration.

The key issue with any of these potential treatments is to balance the oppositional needs of making treatment decisions for

individual patients during epidemic peaks on the basis of clinical studies that involve small numbers of patients with ensuring that well-designed, randomized clinical trials are carried out rapidly to provide proof that they are safe and efficacious. COVID-19 is expected to be active permanently, and several seasons of disease peaks are likely before herd (population) immunity is established. The difficulty is to coordinate rapid

hypothesis-generating studies during this first peak to justify a smaller number of well-controlled large trials to be executed in later peaks to provide the data needed for approval of drugs for COVID-19. Researchers, ethics boards, and regulators are accustomed to developing trial plans over months, not weeks—a time frame that is not afforded during this emergent situation. It is necessary for all involved to work faster and more efficiently and then position the well-justified drugs for registration-enabling trials during the next peak. ■

REFERENCES AND NOTES

1. C. del Rio, P. N. Malani, *JAMA* **323**, 1339 (2020).
2. N. Zhu et al., *N. Engl. J. Med.* **382**, 727 (2020).
3. T. S. Fung, D. X. Liu, *Annu. Rev. Microbiol.* **73**, 529 (2019).
4. G. Li, E. De Clercq, *Nat. Rev. Drug Discov.* **19**, 149 (2020).
5. J. Lan et al., *Nature* **10.1038/s41586-020-2180-5** (2020).
6. V. Monteil et al., *Cell* **10.1016/j.cell.2020.04.004** (2020).
7. M. Hoffmann et al., *Cell* **181**, 271 (2020).
8. M. Wang et al., *Cell Res.* **30**, 269 (2020).
9. M. Mauthe et al., *Autophagy* **14**, 1435 (2018).
10. J. Liu et al., *Cell Discov.* **6**, 16 (2020).
11. M. Renna et al., *J. Clin. Invest.* **121**, 3554 (2011).
12. D. H. Tran et al., *J. Antibiot.* **72**, 759 (2019).
13. Z. Chen et al., *medRxiv*, 2020.03.22.20040758 [Preprint], 10 April 2020. <https://doi.org/10.1101/2020.03.22.20040758>.
14. P. Gautret et al., *Int. J. Antimicrob. Agents* **10.1016/j.ijantimicag.2020.105949** (2020).
15. B. Cao et al., *N. Engl. J. Med.* **10.1056/NEJMoa2001282** (2020).

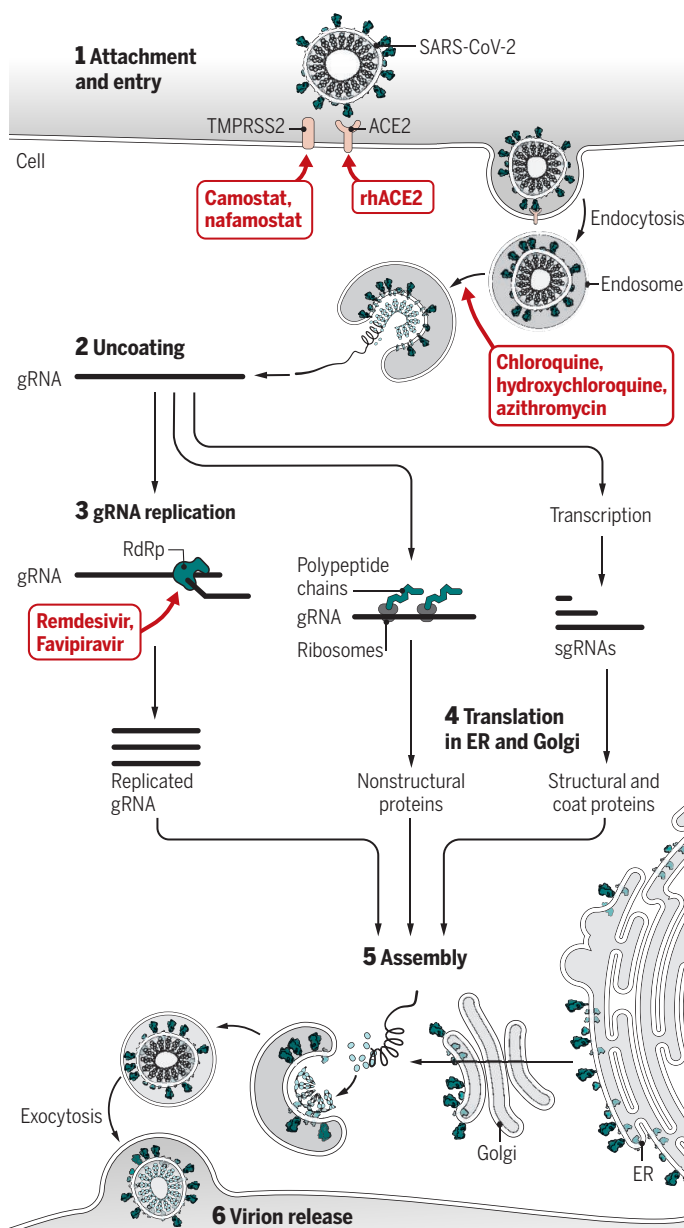
ACKNOWLEDGMENTS

The authors acknowledge the participation of the COVID-19 Unified Research Experts (CURE) Alliance team at the University of Kentucky in discussions underlying this piece. We have no funding supporting the writing of this article and no conflicts of interest. The therapeutics discussed in this Perspective are undergoing clinical testing and are not currently approved for the treatment of COVID-19.

Published online 8 May 2020
10.1126/science.abb9332

Possible targets in the coronavirus life cycle

This simplified coronavirus life cycle shows the processes and proteins that could be therapeutically targeted with existing drugs that have the potential to be repurposed for the treatment of COVID-19.



COVID-19, coronavirus disease 2019; ER, endoplasmic reticulum; gRNA, genomic RNA; RdRp, RNA-dependent RNA polymerase; rhACE2, recombinant human angiotensin-converting enzyme 2; SARS-CoV-2, severe acute respiratory syndrome-coronavirus 2; sgRNA, subgenomic RNA; TMPRSS2, transmembrane protease serine 2.

John Horton Conway (1937–2020)

Innovative mathematician and passionate educator

By **Matt Baker**

John Horton Conway, renowned mathematician of legendary creativity, died on 11 April at age 82. Conway's playful approach to mathematics is visible in his game-changing contributions to a wide variety of mathematical fields. Conway was also a celebrated and gifted educator whose enthusiasm, charisma, and inventiveness captured the public's imagination.

Conway was born in Liverpool, England, on 26 December 1937. He received his mathematics degrees from the University of Cambridge: a bachelor's degree in 1959 and a Ph.D. 5 years later. He was subsequently hired by the university as an assistant lecturer, eventually rising to the rank of professor. In 1987, he accepted a position as the John von Neumann Professor in Applied and Computational Mathematics at Princeton University in Princeton, New Jersey, where he remained for the rest of his career.

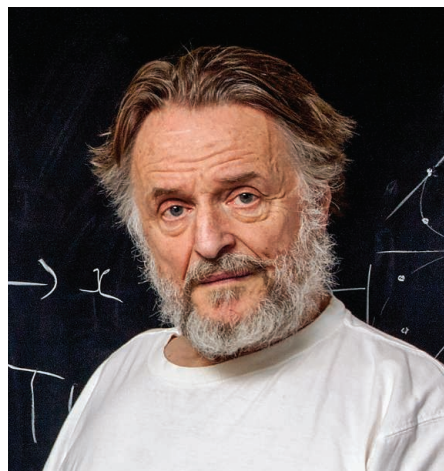
Conway's first big mathematical breakthrough, in 1968, was the determination of the more than eight quintillion symmetries of the Leech lattice, an astonishingly symmetric mathematical structure that governs the tightest and most efficient way to pack spheres together in 24 dimensions. (Think of an array of stacked oranges in a grocery store, except the oranges are 24-dimensional and each one touches exactly 196,560 other oranges.) Through these investigations in the field of group theory, Conway discovered several additional groups, including 3 of the 26 so-called sporadic groups, which play an important role in mathematical physics and the theory of error-correcting codes.

Conway later spearheaded a collaboration that resulted in the *Atlas of Finite Groups*, one of the most important (and longest) books ever written on group theory. Together with *Atlas* coauthor Simon Norton, Conway put forth a series of fascinating conjectures relating the largest of the sporadic groups (which Conway dubbed "The Monster") to the theory of modular forms, a previously unrelated subject arising in complex variables and number theory. These conjectures became known as "Monstrous Moonshine," and Conway's former Ph.D. student Richard

Borcherds won the prestigious Fields Medal for verifying them.

Conway's most famous invention was undoubtedly the Game of Life. An example of what is now called a cellular automaton, this game uses simple rules that lead to wildly unpredictable behavior. A *Scientific American* column written by Conway's lifelong friend and champion Martin Gardner turbocharged the game's popularity and vaulted Conway to international prominence. The simulation became a frequent pastime for budding computer programmers at the dawn of the age of personal computers.

Conway's proudest mathematical moment came about when he parlayed his study of the winning strategies for certain two-player



games into the invention of surreal numbers. These numbers contain not only all real numbers (i.e., infinite decimals such as pi or the square root of 2) but also a staggering cornucopia of new numbers, some infinitely large and some infinitely small. The unification of infinite set theory and the theory of combinatorial games that resulted from this work was hailed by Gardner as "an astonishing feat of legerdemain."

Conway made numerous other contributions to mathematics in a dizzying array of fields. In topology, the Conway polynomial is a fundamental tool for studying knots. In geometry, Conway and mathematician and computer scientist Michael Guy generalized Archimedes' classical enumeration of the 13 three-dimensional so-called "Archimedean solids" to four dimensions. In number theory, Conway and his student William

Schneeberger discovered an important and unexpected result called the "15 theorem" and formulated a conjectural generalization (the "290 theorem"), which was later proved correct. In theoretical physics, Conway and mathematician Simon Kochen proved a striking result in quantum mechanics that they christened the "free will theorem." Conway phrased the result in layman's terms as follows: "If experimenters have free will, then so do elementary particles."

Conway was a captivating teacher. He spent 2 weeks every summer with aspiring young mathematicians at the Canada/USA Mathcamp, where he would give spontaneous lectures on whatever mathematical subject the students requested. His lectures at Cambridge, and later Princeton, were famously idiosyncratic. He would balance objects on his chin, perform magic tricks with a piece of rope, recite the digits of pi (he had memorized more than 1000 of them), and regale students with his astounding knowledge of etymology (a lifelong passion).

I knew Conway through the biennial Gathering 4 Gardner Conference. His notoriously outsized ego was occasionally off-putting, but in the words of biographer Siobhan Roberts, "Conway's is a jocund and playful egomania, sweetened by self-deprecating charm." Once, while visiting Princeton, I ran into Conway, and we ended up chatting for almost 2 hours about different methods for mentally calculating the day of the week for any given date. Conway's astounding speed (in his prime, he could do it in under 2 seconds) made a convincing case for his own doomsday rule, one of his famous contributions to recreational mathematics.

Despite his considerable erudition and numerous mathematical awards, Conway referred to himself as a "professional non-understander," claiming that he only comprehended things after thinking "for ages and [making] them very, very simple." He followed his insatiably peripatetic curiosity no matter its direction, and the professional success he achieved in this way led him to adopt the dictum: "Thou shalt stop worrying and feeling guilty; thou shalt do whatever thou pleasest." Another of Conway's guiding principles was to always go multiple steps beyond what any reasonable person would do.

In addition to his scientific breakthroughs, Conway transformed the public perception of mathematics. His singularly creative ideas have inspired generations of mathematicians. Although "recreational mathematics" may seem like an oxymoron to some, the term perfectly captures Conway's joyful approach to both serious and recreational mathematical discoveries. ■

School of Mathematics, Georgia Institute of Technology, Atlanta, GA, USA. Email: mbaker@math.gatech.edu

10.1126/science.abc5331

POLICY FORUM

RESEARCH ETHICS: COVID-19

Ethics of controlled human infection to address COVID-19

High social value is fundamental to justifying these studies

By Seema K. Shah, Franklin G. Miller, Thomas C. Darton, Devan Duenas, Claudia Emerson, Holly Fernandez Lynch, Euzebiusz Jamrozik, Nancy S. Jecker, Dorcas Kamuya, Melissa Kapulu, Jonathan Kimmelman, Douglas MacKay, Matthew J. Memoli, Sean C. Murphy, Ricardo Palacios, Thomas L. Richie, Meta Roestenberg, Abha Saxena, Katherine Saylor, Michael J. Selgelid, Vina Vaswani, Annette Rid

Development of an effective vaccine is the clearest path to controlling the coronavirus disease 2019 (COVID-19) pandemic. To accelerate vaccine development, some researchers are pursuing, and thousands of people have expressed interest in participating in, controlled human infection studies (CHIs) with severe acute respiratory syndrome–coronavirus 2 (SARS-CoV-2) (1, 2). In CHIs, a small number of participants are deliberately exposed to a pathogen to study infection and gather preliminary efficacy data on experimental vaccines or treatments. We have been developing a comprehensive, state-of-the-art ethical framework for CHIs that emphasizes their social value as fundamental to justifying these studies. The ethics of CHIs in general are underexplored (3, 4), and ethical examinations of SARS-CoV-2 CHIs have largely focused on whether the risks are acceptable and participants could give valid informed consent (1). The high social value of such CHIs has generally been assumed. Based on our framework, we agree on the ethical conditions for conducting SARS-CoV-2 CHIs (see the table). We differ on whether the social value of such CHIs is sufficient to justify the risks at present, given uncertainty about both in a rapidly evolving situation; yet we see none of our disagreements as insurmountable. We provide ethical guidance for research sponsors, communities, participants, and the essential independent reviewers considering SARS-CoV-2 CHIs.

Author affiliations are listed in the supplementary materials. Email: seema.shah@northwestern.edu

SUFFICIENT SOCIAL VALUE

CHIs have a long, complicated history. They have contributed to substantial improvements in clinical and public health practice, including the recent licensure of two vaccines (5), but also involved some unethical research (3). The first step in justifying SARS-CoV-2 CHIs, especially as they would involve major uncertainty and controversy, is to demonstrate their high social value. Crucially, SARS-CoV-2 CHIs should address relevant, unresolved scientific questions in rigorously designed and conducted experiments.

SARS-CoV-2 CHIs could have high social value in several ways. For example, they could help prioritize among the almost 100 investigational vaccines and over 100 experimental treatments for COVID-19 currently in development. CHIs could help identify the most promising agents, which would inform the design of larger trials, guide decisions to scale up manufacturing early, and thereby accelerate product development and implementation. If they saved even a few months of vaccine development (1), SARS-CoV-2 CHIs would contribute to faster control of the pandemic and reduce the need for, and associated costs of, physical distancing measures, providing substantial benefits for much of the world's population (including the most vulnerable).

To achieve high social value in this way, coordination of stakeholders is essential. Sponsors of SARS-CoV-2 CHIs should delineate a credible path forward from CHIs to rigorous field studies, and eventually toward scaled-up production. This is a considerable challenge given the rapidly evolving research response to the pandemic; many approaches to accelerating product development are already appropriately being pursued in parallel. It is therefore essential to plan and evaluate SARS-CoV-2 CHIs as a complement, not an alternative, to these other approaches and ensure that CHI results are integrated into the dynamic COVID-19 research landscape. For example, the World Health Organization is convening sponsors of SARS-CoV-2 CHIs to increase

transparency and promote coordination. Research sponsors should lead by establishing and enforcing standards for rapid data collection, dissemination, and sharing that permit aggregation of results across CHIs. Medical journals should require compliance with these standards before accepting manuscripts. Regulatory agencies should collaborate with sponsors, researchers, and policy-makers to define how CHI data will inform or modify larger trials, licensure, and manufacturing. Finally, sponsors and governments should implement mechanisms to ensure widespread, equitable access to proven products whose development was accelerated by SARS-CoV-2 CHIs. Such wide-ranging stakeholder coordination is difficult but important to demonstrate high social value. Though not achieved for proposed Zika virus CHIs during the 2015–2016 epidemic, it did occur later (6).

SARS-CoV-2 CHIs could have high social value in other ways, and individual CHIs could address multiple scientific questions. For example, CHIs could clarify dynamics of infection, viral pathogenesis, and risk of vaccine pathogenesis or identify correlates of protection—all of which could inform the development and implementation of vaccines. CHIs could also illuminate poorly understood parameters for modeling the pandemic and public health responses, including who is infectious and when and how infections occurred. This information is difficult to collect by observation alone, and existing animal models do not fully replicate clinical disease seen in humans. Additionally, if the pandemic wanes before larger trials are completed, SARS-CoV-2 CHIs could be critical for advancing research until the next outbreak, as with Zika virus (6). All of these paths to high social value would require similar, extensive coordination with relevant stakeholders.

SARS-CoV-2 CHIs admittedly have limited generalizability, as they would need to be conducted with low-risk populations (see below) with a non-natural mode of infection. Therefore, although some propose replacing efficacy trials with SARS-CoV-2 CHIs (1), it is more likely that CHIs accelerate vaccine or treatment development by informing larger trials, not by making such trials redundant. Yet almost all disease models or trial designs require some extrapolation or further testing. For example, field trials with frontline workers could also accelerate vaccine development, but they would not include older, retired individuals.

Thus, there are many potential ways in which SARS-CoV-2 CHIs could have high social value. Before their initiation, it is essential that the given social value is judged as compelling enough to justify its pursuit.

Ethical framework for SARS-CoV-2 controlled human infection studies (CHIs)

SUFFICIENT SOCIAL VALUE

Identify and address relevant, unresolved scientific questions in rigorously designed and conducted experiments

- Use rigorous methods to develop CHI models, including high-quality manufacturing and process of challenge strains
- Define and regularly review priority scientific questions, e.g., selecting the most promising vaccine and treatment candidates; identifying correlates of protection; clarifying infection dynamics, mechanisms of disease, and possible vaccine pathogenesis
- Coordinate with stakeholders to ensure SARS-CoV-2 CHI results will affect future research, clinical, or public health practice and delineate a credible path forward for CHI results to make an impact
- Establish and enforce standards for data collection in SARS-CoV-2 CHIs
- Share data, samples, and challenge strains appropriately
- Disseminate SARS-CoV-2 CHI results quickly through open-access publication

Realize benefits by facilitating equitable access to proven safe and effective products

- Use mechanisms such as compulsory licensure under Trade-Related Aspects of Intellectual Property Rights (TRIPS) agreement, march-in rights against patents under U.S. Bayh-Dole Act, U.S. Food and Drug Administration's priority review voucher program

REASONABLE RISK-BENEFIT PROFILE

Identify and reduce risks

- Enroll younger, adult participants without comorbidities; refine and update eligibility criteria in light of new evidence
- Monitor closely; provide prompt, free treatment and compensation for research-related injury
- Confine participants in in-patient isolation for at least 14 days
- Inform public health officials about study in advance

Ensure reduced risks are ethically acceptable

- Risks should not exceed upper limits
- Risks should be reasonable in relation to social value

CONTEXT-SPECIFIC STAKEHOLDER ENGAGEMENT

Engage public

- Create community advisory boards; use media to inform and engage
- Gather public input through informed opinion surveys
- Adapt engagement strategies to physical distancing as needed

Coordinate with international research, clinical, public health community

- Engage researchers, sponsors, regulators, ministries of health, etc. before, during, and after study implementation

SUITABLE SITE SELECTION

Consider feasibility of recruitment, risk, generalizability, availability of infrastructure, potential effects on local health care system

- Select location with available expertise
- Bring in extra resources so as not to unduly compromise pandemic response

FAIR PARTICIPANT SELECTION

Enroll low-risk groups with capacity to provide voluntary informed consent

- Enroll younger, adult participants without comorbidities

ROBUST INFORMED CONSENT

Ensure participant understanding

- Use evidence-based, context-specific consent materials
- Test participants on understanding of key criteria: deliberate infection; risks and burdens; potential social value; study purpose; uncertainty; restrictions on liberty to protect others

PROPORTIONATE PAYMENT

Avoid undue influence, exploitation, incentives to withhold information

- Compensate participants for their time to avoid exploitation and inequities in access to CHIs
- Set objective, verifiable eligibility criteria (in case money tempts participants to withhold disqualifying information)

REASONABLE RISK-BENEFIT PROFILE

For SARS-CoV-2 CHIs to be ethically permissible, risks to participants, study personnel, and third parties should be minimized, reasonable in relation to the social value of the research, and below the upper limits of acceptable risk (7, 8). There are both scientific unknowns about SARS-CoV-2 and moral disagreements about upper limits to risk. Although research inherently involves uncertainty, this situation warrants a cautious approach to evaluating SARS-CoV-2 CHIs and revisiting risk/benefit judgments as new evidence emerges.

Risk minimization should focus primarily on reducing the likelihood of serious and irreversible harms. To minimize risks to participants, SARS-CoV-2 CHIs should recruit young people without underlying medical

conditions who face lower mortality risks from COVID-19 (9). Key uncertainties remain regarding other potentially serious and irreversible harms of SARS-CoV-2 infection in young people, such as cardiac or neurological injury. Accordingly, long-term follow-up of CHI participants is critical.

To minimize risks to study personnel, participants should be in inpatient isolation, with contact reduced to the extent possible and robust personal protective equipment provided. Both participants and personnel should be carefully monitored, promptly managed when symptomatic, and provided any proven targeted treatments or offered enrollment into appropriate clinical trials. To minimize risks to third parties outside the research, researchers should notify public health authorities about the studies in

advance and ensure that participants who withdraw take appropriate precautions to avoid spread to others.

Participants might benefit from controlled infection and/or vaccination if they become immune to SARS-CoV-2. However, the degree and duration of naturally acquired and vaccine-derived immunity are currently unknown. Some participants would also receive placebo vaccines, and most investigational vaccines prove ineffective. The potential benefits of participation thus should be given little, if any, weight. Instead, risks to participant should be justified by the social value of SARS-CoV-2 CHIs, with higher risks requiring higher social value.

Even when research has high social value and involves competent consenting

adults, there is substantial consensus that risks to participants should not exceed an absolute upper limit. Regulations and ethics guidance do not clearly delineate this limit. Some commentators have argued that it should not exceed a 1% risk of death or the risks posed by activities that, like research, expose some people to risk to benefit others, such as living organ donation (8, 10). Although these are imperfect analogies to research, they provide helpful context for evaluating limits of acceptable research risk.

Current data on SARS-CoV-2 infection come from relatively small samples with missing data points and are still being scrutinized. Data suggest that 20- to 44-year-olds with diagnosed infection—including those with underlying conditions—have a mortality risk less than 0.2% (11). But diagnostic testing has been limited, making the number of undiagnosed infections unknown. One attempt to account for these limitations estimates that healthy adults aged 20 to 29 have a 0.03% risk of death and a 1.1% risk of hospitalizations (9). These risks could be further reduced by refining eligibility criteria based on emerging data. Recognizing the uncertainties, risks from SARS-CoV-2 CHIs appear comparable to the risks from some other research and activities similar to research (table S1). They also seem to fall below the upper risk limits proposed for research. For third parties who could be exposed to infection from CHI participants, there is no consensus on what level of risk is acceptable (12); however, with the above safeguards, these risks could be minimized to be negligible.

CONTEXT-SPECIFIC STAKEHOLDER ENGAGEMENT

CHIs have a checkered history (3), and it can be counterintuitive for the public that researchers would infect people with disease-causing pathogens. Although the current pandemic context with widespread physical distancing might complicate public engagement, it remains important and feasible as SARS-CoV-2 CHIs are developed. For example, public opinion surveys could identify concerns and information deficits, and researchers could engage the media or convene virtual advisory groups. Maintaining transparency and accountability to diverse communities is important for mitigating potential mistrust, especially in a pandemic (13). As noted above, engagement with stakeholders in the research community, health professionals, and policy-makers is also critical for ensuring that the results from SARS-CoV-2 CHIs translate into social benefits.

SUITABLE SITE SELECTION

Selecting suitable sites for SARS-CoV-2 CHIs requires considering risks to participants, study personnel, and third parties; feasibility of recruitment; availability of necessary infrastructure; and potential effects on local pandemic responses. Sites should be selected for sound scientific reasons while avoiding especially vulnerable populations. For example, performing CHIs in locations with high community spread of SARS-CoV-2 could be an acceptable way to reduce relative risks for participants, provided that high transmission is not due to underlying injustices. Given that participants would require testing, medical attention, and treatment, and research personnel would require personal protective equipment, sponsors should also demonstrate to ethics review boards or public health authorities that CHIs will not unduly compete for scarce resources and thereby compromise the local pandemic response. All sites should have sufficient capacity to conduct rigorous studies, provide high-quality care to participants, and minimize research risks. Sites experienced with conducting CHIs might be favored to ensure that studies and local public engagement can be launched quickly, effectively, and responsibly.

FAIR PARTICIPANT SELECTION

Selecting participants fairly for SARS-CoV-2 CHIs primarily requires considering fair distribution of research risks and burdens. Because of the uncertainty and potential high risk involved, participants who are at relatively low risk of serious and irreversible harm and have capacity to give their own consent should be selected (i.e., young, healthy and competent adults).

ROBUST INFORMED CONSENT

There is widespread consensus on obtaining high-quality informed consent for CHIs and using rigorous procedures to maximize participant understanding. Evidence-based approaches to consent include requiring participants to pass a test on key study information (14). Ongoing informed consent will be important as new data emerge, notably on the risks of SARS-CoV-2 infection.

PROPORTIONATE PAYMENT

Members of our group disagree about the ethical permissibility of offering payment to CHI participants, and there may be relevant regulatory limits in different jurisdictions. Nevertheless, as SARS-CoV-2 CHIs require confinement and follow-up, fairness seems to demand offering participants compensation for their time. This may total several thousand dollars in the United States, as-

suming compensation at a fair minimum wage for unskilled labor, as in other CHIs. By contrast, incentives beyond compensation could be avoided, given the number of people already indicating willingness to participate. Concerns that the undue influence of monetary compensation compromises risk judgments are unsupported by the available data, as financial motivations are associated with greater attention to risk (15). Moreover, a rigorous informed consent process could maximize understanding. In case payment tempts participants to withhold disqualifying information, eligibility criteria should be objectively verifiable.

CONCLUSION

Given the extraordinary nature of the pandemic, our framework and analysis support laying the groundwork for SARS-CoV-2 CHIs—for example, by developing a challenge strain, drafting consensus protocols that address ethical concerns, and engaging stakeholders to enhance their social value, minimize risks, and build public trust. ■

REFERENCES AND NOTES

1. N. Eyal, M. Lipsitch, P. G. Smith, Human challenge studies to accelerate coronavirus vaccine licensure. *J. Infect. Dis.* jiaa152 (2020).
2. www.thecovidchallenge.org/
3. B. Bamberg, M. Selgelid, C. Weijer, J. Savulescu, A. J. Pollard, *Public Health Ethics* **9**, 92 (2016).
4. E. Jamrozik, M. J. Selgelid, *Human Challenge Studies in Endemic Settings: Ethical and Regulatory Issues*, Springer Briefs in Ethics (Springer, 2020).
5. M. Roestenberg, M.-A. Hoogerwerf, D. M. Ferreira, B. Mordmüller, M. Yazdanbakhsh, *Lancet Infect. Dis.* **18**, e312 (2018).
6. K. S. Vannice *et al.*, *Vaccine* **37**, 863 (2019).
7. A. Rid, D. Wendler, *Kennedy Inst. Ethics J.* **21**, 141 (2011).
8. F. G. Miller, S. Joffe, *J. Med. Ethics* **35**, 445 (2009).
9. R. Verity *et al.*, *Lancet Infect. Dis.* **10**, 1016/S1473-3099(20)30243-7 (2020).
10. D. B. Resnik, *Theor. Med. Bioeth.* **33**, 137 (2012).
11. S. Bialek *et al.*, *MMWR Morb. Mortal. Wkly. Rep.* **69**, 343 (2020).
12. S. K. Shah *et al.*, *Science* **360**, 158 (2018).
13. AVAC, Good Participatory Practice Guidelines for HIV Prevention Research, www.avac.org/resource/good-participatory-practice-guidelines-biomedical-hiv-prevention-trials-second-edition.
14. A. Nishimura *et al.*, *BMC Med. Ethics* **14**, 28 (2013).
15. L. Stunkel, C. Grady, *Contemp. Clin. Trials* **32**, 342 (2011).

ACKNOWLEDGMENTS

The opinions expressed in the article are the authors' and do not reflect the views of organizations with which the authors have affiliations, including the National Institutes of Health, the Department of Health and Human Services, or the United States government. This work was primarily supported by a Making a Difference Grant from the Greenwall Foundation (S.K.S., A.R., R.P., D.D.), along with support from the Wellcome Trust (S.K.S., E.J., D.K., M.K., R.P., M.J.S., V.V.), Brocher Foundation (S.K.S., A.R., R.P., D.D., T.C.D., H.F.L., E.J., N.S.J., D.K., J.K., D.M., S.C.M., T.L.R., M.R., A.S., M.J.S., V.V.), and NIH Clinical Center Department of Bioethics (A.R.). The authors also thank C. Chui, K. Littler, P. Pitisuttithum, and M. Yu for their contributions, and M. Danis, C. Grady, M. Nicolini, J. Ochoa, and H. Taylor for helpful discussion.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6493/832/suppl/DC1

Published online 7 May 2020

10.1126/science.abc1076

sciencemag.org **SCIENCE**



BOOKS *et al.*

ENTOMOLOGY

Butterflies and the people who love them

A meandering investigation hints at how much is left to learn about these charismatic insects

By **Anurag A. Agrawal**

The idea of a life cycle is ubiquitous, from industrial resource extraction, production, consumption, and disposal to the various stages through which biological entities pass. And yet there was a time, not so long ago, when the concept of a life cycle was foreign. Imagine 17th-century observers of the Lepidoptera, watching them move from egg to caterpillar to butterfly. What did they make of these creatures? “Are they two species or three?” they might have wondered.

In *The Language of Butterflies*, Wendy Williams chronicles some of the key events in the history of butterflies, spanning the geologic record to current population declines. The book flits from personal journey, to the work of scientists, to the biology of butterflies, weaving a conversational and accessible lyric. The target audience is interested naturalists, butterfly lovers, and science enthusiasts who want to know more about the lives of butterflies and those who chase them.

At her best, Williams digs deeply into the lives of both butterflies and scientists—reporting, for example, on the spectacular discoveries, personal life, and writings of Maria Sibylla Merian. One of the great naturalists of the 17th century, Merian directly observed insects in Europe and Suriname, beautifully rendered their life cycles, and connected the dots between egg, caterpillar, and butterfly in insightful books.

The reviewer is at the Department of Ecology and Evolutionary Biology and the Department of Entomology, Cornell University, Ithaca, NY 14853, USA, and is the author of *Monarchs and Milkweed* (Princeton Univ. Press, 2017). Email: agrawal@cornell.edu

Williams’s treatment of the accumulation of knowledge about butterflies through history—from Merian’s journey to the current state of biological knowledge—is informative and illuminating.

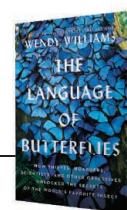
Although *The Language of Butterflies* frequently shifts topics and lacks a thesis or core message, several threads bear notice. For example, the book advances the notion that there is something special about color being eye-catching to many animals. Unfortunately, Williams fails to offer insight into the evolution of color, its diversity, and its impact on ecological interactions. One is left wondering how much of a butterfly’s coloration is dictated by the mating benefits it confers, advertisement of toxicity to their predators, and evolutionary pressure to absorb particular wavelengths of light. Do the contributions of these pressures vary among species?

Midway through the book, the focus turns to one of the world’s most popular insects, the monarch butterfly. Williams narrows in on the smaller California population of monarchs, which often receives less attention than the hordes in the east that travel to Mexico. Here, we meet schoolchildren and entomologists and learn about the coevolutionary interaction between monarchs and milkweed, the species’ only larval food source.

Sometimes Williams’s poetic license obscures her point, as when she discusses the evolution of the Lepidoptera: “When flowers evolved, they gradually enslaved some of the moths and turned them into butterflies, who would perform important duties for their flower masters.” Even so, as she reveals, there is still much to learn about how species such as the monarch

The Language of Butterflies

Wendy Williams
Simon and Schuster, 2020.
240 pp.



transform from egg to caterpillar to butterfly, about the nature of their long-distance migration, and about the causes of their population declines in both eastern and western North America.

Although an exclusive focus on monarchs would be easy to defend, Williams also writes about conservation efforts for lesser-known butterflies—for example, the Fender’s blue butterfly, which relies on a rare lupine to complete its breeding cycle. The Fender’s blue conservation project in Oregon mirrored an earlier conservation success, that of the endangered large blue in Europe. In both cases, Williams documents how unraveling the basic biology and entire life cycle was critical to bringing these butterflies back.

Still, I would have liked to have seen stronger comparisons across different species. How do the ecologies of monarchs and Fender’s blue butterflies differ, for example, and what does this mean with regard to each species’ role as a conservation case study? What have we learned about evolution from the divergent strategies that different butterfly lineages have taken?

While *The Language of Butterflies* unlocks no intellectual puzzles, it does extend the common view of butterflies beyond mere eye candy to connect readers to nature and biology in a potentially generative manner. ■

10.1126/science.abb3531

ECOLOGY

Diversity and the dinner plate

An intimate exploration of foraged flavors brings nuance to the wild food discussion

By **Lenore Newman**

For those of us who travel for a living, the coronavirus disease 2019 (COVID-19) crisis carries extra bite. Books such as Gina Rae La Cerva's *Feasting Wild* offer a balm of sorts, transporting the reader out of lockdown and into a world of little-known and endangered flavors on the margins of where human society rubs against wild ecosystems.

Feasting Wild is an ambitious book that asks a big question: What does it mean to eat wild food? Like most big questions, the answer can differ depending on the scale. La Cerva approaches this question both from a personal perspective and through the voices of diverse groups, ranging from Polish foresters to urban chefs foraging for wild treats in graveyards. She shines light on the larger context of humanity's relationship with wild foods in snippets and glances, giving the reader little jewels of introspection and observation. This is a journalistic exploration based on phenomenological research—and a very good one.

La Cerva warmly welcomes readers along for an intimate journey into the world of wild food. Together, we visit corners of the world's food system that have rarely been explored in print at this level of detail—places such as Borneo, where La Cerva describes the exacting preparation of a bird's nest for consumption, and the forests of Sweden, where she takes us along on a moose hunt and then walks us through the difficult work of processing the meat.

We also vicariously experience some very interesting meals. At Noma, a high-end, avant-garde restaurant in Copenhagen, for example, unusual ferments ranging from lacto-fermented raspberries to preserved mosses offer a good bridge into the complicated world of bushmeat preparation.

At times, the intimacy of the book crosses into the literal. There is a love affair and

hints of a free-range childhood, seasoned with sketches any field researcher would instantly recognize: humidity, uneasy interviews, rattling bush planes, and questionable motorcycle trips. La Cerva reveals the landscape in brightly lit detail and gives generously of herself, and the result makes for a suitably satisfying feast.

The opening third of the book jumps rapidly from Copenhagen to the forests of Poland and includes glimpses of La Cerva's childhood in the desert of New Mexico. The richness of the book shows through in glimmers here, but I felt the overall theme of wild food, loss, and change was scattered at first. Some of this was personal.



Quail eggs at Noma evoke the wild foods for which the restaurant is known.

I grew up in a town similar to the Danish commune of Christiania; the “free society” that La Cerva marvels at while exploring Copenhagen reminds me of bickering over chores, and her sense of wonder falls flat.

La Cerva finds her voice in Part Two. Her description of the bushmeat trade in the Democratic Republic of Congo is one of the most tightly crafted and engaging pieces of writing I have enjoyed in years. Here, the pages flow by like the great rivers of the Congo Basin, taking the reader through the realities of the bushmeat trade and the desires of the eater. La Cerva is not blind to the pressure that bushmeat consumption puts on endangered species or to the economic realities that drive the bushmeat trade. However, she also captures the cultural importance of bushmeat, a factor

Feasting Wild: In Search of the Last Untamed Food

Gina Rae La Cerva
Greystone Books, 2020.
336 pp.



we will need to consider as conversations regarding the role of wild foods in the spread of zoonotic viruses continue.

We follow the meat as it travels to the markets of Paris and, in doing so, fall deeper into the author's life. Her love affair with the man we know only as “the Hunter” and the wildness he represents provide a counterpoint for the wild foods she seeks to capture.

The final third of the book is a fitting denouement. In the woods of Sweden, wood smoke and the chill of the passing seasons help La Cerva tie up open narratives. Her sudden shift in this section to the exploration of the edible bird's nest industry came as a desert of sorts. Although she does not manage to access the wild bird caves where edible nests are harvested, she does uncover the reality of nest farming. When she tastes tea brewed from the nests and describes its underwhelming flavor, we are reminded that in the case of wild food, the myth often overshadows the tepid reality. I was a little disappointed that this section was so short, but that is my usual reaction to desserts, both real and literary.

Feasting Wild did have one weakness that is worth mentioning. The book cleaves tightly to the narrative of wilderness as paradise lost, a fallen kingdom. As William Cronon noted in 1995 in the *New York Times Magazine*, prompting a canon of related literature, wilderness is a profoundly human creation, the result of a particular set of human cultures at a specific moment in their history (*1*). This is, however, a minor quibble, given how deeply this book spoke to me as a kindred traveler. ■

REFERENCES AND NOTES

1. W. Cronon, “The trouble with wilderness,” *The New York Times Magazine*, 13 August 1995.

The reviewer is at the Department of Geography and the Environment, University of the Fraser Valley, Abbotsford, BC, Canada, and the author of *Lost Feast: Culinary Extinction and the Future of Food* (ECW Press, 2019). Email: lenore.newman@ufv.ca

10.1126/science.abb8580



The rare Chinese box turtle (*Cuora flavomarginata*) continues to be traded on the black market.

Edited by Jennifer Sills

China's wild turtles at risk of extinction

China's turtle species diversity ranks third in the world, with 34 species including 3 tortoises, 5 sea turtles, and 26 freshwater turtles (1). However, with the exception of three species that lack sufficient data, all of China's turtle species have been listed as Endangered or Critically Endangered on the Red List of China's Vertebrates, making them the most endangered group of vertebrates in the country (1). Evidence suggests that most wild populations of turtles in China declined by more than 90% between 1980 and 2012 (2). If China does not take steps to mitigate the risks to wild turtles, they could face imminent extinction.

The catastrophic decrease of the wild turtle population is primarily the result of overexploitation (2, 3). Because turtles symbolize luck and longevity in Chinese culture, they are widely kept as pets and used for food and traditional medicine (3, 4). To meet the huge consumption demand, the turtle farming industry has developed rapidly in China since the 1990s (5). Farmers purchase wild-caught turtles to improve the reproductive capability of farmed stocks (6). Some rare species, such as box turtles (genus *Cuora*), are very expensive in the black market (7), and high profits have stimulated poaching of wild turtles (2).

Wild turtles in China face other threats

as well. With the rapid economic development in recent decades, construction of hydropower plants, coastal development, water pollution, and other human activities have destroyed the turtles' habitat (8, 9). Meanwhile, the invasion of alien turtles threatens the survival of native turtles (10). For example, the red-eared slider (*Trachemys scripta elegans*), one of the world's 100 most invasive species, is farmed in China and has been widely found in the wild, causing serious ecological risks (10).

The Chinese government must urgently implement measures to protect wild turtles. Ecological education efforts should work to shift the expression of cultural love for turtles from consumption to conservation. The government should also strengthen the enforcement of the Wildlife Protection Law of China (11) to crack down on poaching and illegal trade of wild turtles and to prevent and control the invasion risk of alien turtles. Finally, despite their endangered status on the Red List (1), only one-third of China's wild turtles are currently included on the National Key-Protected Species List (12), an appendix of the Wildlife Protection Law of China that lists species in need of protection. To ensure legal and enforceable protection, China should update the Protected Species List to include all endangered turtle species.

Jun Wu¹, Yanqing Wu¹, Dingqi Rao², Ting Zhou³,
Shiping Gong^{4*}

¹Key Laboratory of Biosafety, Nanjing Institute of Environmental Sciences, Ministry of Ecology

and Environment, Nanjing 210042, China.

²State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Science, Kunming 650223, China. ³Hainan Academy of Forestry, Haikou 571100, China. ⁴Guangdong Key Laboratory of Animal Conservation and Resource Utilization, Guangdong Public Laboratory of Wild Animal Conservation and Utilization, Guangdong Institute of Applied Biological Resources, Guangdong Academy of Sciences, Guangzhou 510260, China.

*Corresponding author. Email: gsp621@163.com

REFERENCES AND NOTES

1. Z. G. Jiang *et al.*, *Biodivers. Sci.* **24**, 500 (2016) [in Chinese].
2. S. P. Gong *et al.*, *Curr. Biol.* **27**, 170 (2017).
3. Z. H. Zhou, Z. G. Jiang, *Chelon. Conserv. Biol.* **7**, 28 (2008).
4. S. M. Cheung, D. Dudgeon, *Aquat. Conserv. Mar. Freshw. Ecosyst.* **16**, 751 (2006).
5. J. Z. Ma *et al.*, "Report on sustainable development strategy of China's wildlife farming industry" (Consulting Research Project of Chinese Academy of Engineering, 2017) [in Chinese].
6. H. T. Shi, J. F. Parham, M. Lau, T. H. Chen, *Conserv. Biol.* **21**, 5 (2007).
7. D. Gaillard, L. Lin, H. T. Shi, S. J. Luo, *Herpetol. Conserv. Biol.* **12**, 33 (2017).
8. X. Y. Hong *et al.*, *Chelon. Conserv. Biol.* **18**, 68 (2019).
9. J. Wang *et al.*, *Chin. J. Wildl.* **40**, 1070 (2019) [in Chinese].
10. S. P. Gong, J. B. Yang, Y. Ge, D. Gaillard, *Chin. J. Wildl.* **39**, 373 (2018) [in Chinese].
11. The National People's Congress of the People's Republic of China, "The Wildlife Protection Law of China" (2018); www.npc.gov.cn/npc/c12435/201811/f4d2b7a-3024b41ee8ea0ce54ac117daa.shtml [in Chinese].
12. The Department of Forestry and Grassland, "The National Key-protected Species List" (2018); www.forestry.gov.cn/main/3954/content-1063883.html [in Chinese].

10.1126/science.abc0997

COVID-19 recovery can benefit biodiversity

Coronavirus disease 2019 (COVID-19) is a global crisis. Severe interruptions to international trade and travel are crippling economies and forcing reevaluation of economic, health, and environmental trajectories. Given that COVID-19 has triggered widespread changes in human behavior and reductions in pollution (1, 2), it presents opportunities for further positive change. Lockdowns have spurred households to rethink consumer needs, making now an opportune time to promote sustainable consumer choices that will become more engrained with prolonged exposure (1). How we emerge from the state of lockdowns will drive a new world economy with lasting effects on global biodiversity and supply chains (3, 4).

The COVID-19 pandemic has the potential to trigger enormous effects on biodiversity and conservation outcomes. This virus emerged due to wildlife exploitation (5), and the risk of new diseases increases with environmental degradation (6). Past events such as pandemics, wars,

and financial crises have also triggered quantifiable environmental changes (7, 8). We can learn from such events to guide effective conservation strategy. National governments and intergovernmental organizations should adopt clear strategies to safeguard both biodiversity and human health throughout the COVID-19 recovery.

Active promotion and implementation of certain strategies could tip the balance in favor of positive biodiversity outcomes. We can reboot economies while protecting humans and nature by redesigning trade networks and supply chains to localize and better support sustainable consumer options. We can also strengthen environmental protections, improve environmental monitoring through better use of automation, and ensure that conservation funding schemes remain active.

Environmental policy has already moved in both directions. Although in some places, environmental protections have weakened (9), in others, governments have banned animal trade (3, 10) and aim to localize supply chains to increase resource security (11). Blanket wildlife trade bans are not the answer (3), but appropriately nuanced strategies that incorporate such measures should be encouraged. As we progress into a post-COVID-19 world, recovery strategies can be optimized to benefit biodiversity conservation and protect human health.

Ryan M. Pearson*, Michael Sievers, Eva C. McClure, Mischa P. Turschwell, Rod M. Connolly
Australian Rivers Institute—Coast & Estuaries, School of Environment and Science, Griffith University, Gold Coast, QLD 4222, Australia.
*Corresponding author.
Email: r.pearson@griffith.edu.au

REFERENCES AND NOTES

1. M. J. Cohen, *Sustain. Sci. Pract. Pol.* **16**, 1 (2020).
2. D. T. Molintas, "Analysis of Coronavirus and carbon emissions," MPRA Paper 98858 (University Library of Munich, Germany, 2020).
3. H. Wang *et al.*, *Science* **367**, 1435 (2020).
4. H. Zhao, *Science* **367**, 1436 (2020).
5. P. Zhou *et al.*, *Nature* **579**, 270 (2020).
6. F. Keesing *et al.*, *Nature* **468**, 647 (2010).
7. J. Pongratz *et al.*, *Holocene* **21**, 843 (2011).
8. J. Sayer *et al.*, *Intl. For. Rev.* **14**, 90 (2012).
9. Amnesty International, "USA: Immediately revoke COVID-19 suspension of environmental protections," *Amnesty International* (2020).
10. N. Yang, P. Liu, W. Li, L. Zhang, *Science* **367**, 1434 (2020).
11. M. Foley, J. Duke, "Coronavirus triggers Australian self-sufficiency push," *The Sydney Morning Herald* (2020).

10.1126/science.abc1430

COVID-19 spotlights medical diagnostics

The coronavirus disease 2019 (COVID-19) pandemic highlights the importance of the field of medical diagnostics.

Governments are trying to avert crisis conditions by opening makeshift testing units and recruiting nonclinical research staff to conduct testing (1), but this strategy is not a long-term solution. To increase the number of medical diagnosticians, this career path should be encouraged, valued, and adequately funded. Diagnostic expertise will likely become even more vital as our rapidly aging societies continue to challenge a strained health care system (2, 3).

Although constant steps are undertaken to improve the working conditions of doctors, paramedics, and nurses [e.g., (4, 5)] as well as to promote these career choices among the young generations (6, 7), the field of medical diagnostics lags (6). Medical universities focused on educating first-line medical staff often give nonclinical degree programs lower priority and funding. Medical diagnostics graduates, saddled with the less prestigious perception of their profession, rarely request higher wages or better working conditions (6). Diagnostics facilities play an important role in the functioning of health care in both everyday and emergency situations, and yet they are often overlooked in budget plans (8–10). Ensuring proper training, funding, and esteem for diagnostic personnel and facilities is crucial to a successful health care system.

Maurycy Jankowski¹, Paul Mozdziak², Bartosz Kempisty^{1,3,4*}

¹Department of Anatomy, Poznan University of Medical Sciences, Poznan, 60-781 Poland.
²Graduate Physiology Program, North Carolina State University, Raleigh, NC 27695-7608, USA.
³Department of Histology and Embryology, Poznan University of Medical Sciences, 60-781, Poznan, Poland.
⁴Department of Veterinary Surgery, Institute of Veterinary Medicine, Nicolaus Copernicus University in Toruń, 87-100 Toruń, Poland.
*Corresponding author.
Email: bkempisty@ump.edu.pl

REFERENCES AND NOTES

1. J. M. Sharfstein, S. J. Becker, M. M. Mello, *JAMA* **10.1001/jama.2020.3864** (2020).
2. M. J. Binnicker, *Clin. Chem.* **10.1093/CLINCHEM/HVAA071** (2020).
3. Y. M. Arabi, S. Murthy, S. Webb, *Intensive Care Med.* **10.1007/s00134-020-05955-1** (2020).
4. "Emergency medical workers deserve pay equity," *The New York Times* (2019).
5. Department of Health and Social Care, "Pay rise announced for thousands working in medicine," gov.uk (2019).
6. American Society for Clinical Laboratory Science (ASCLS), "Addressing the clinical laboratory workforce shortage" (ASCLS, 2018).
7. I. Kagan *et al.*, *Int. Nurs. Rev.* **62**, 368 (2015).
8. A. Remuzzi, G. Remuzzi, *Lancet* **10.1016/s0140-6736(20)30627-9** (2020).
9. J. Hopman, B. Allegranzi, S. Mehtar, *JAMA* **10.1001/jama.2020.4169** (2020).
10. H. Kuchler, "Coronavirus testing shortages: What's the problem?," *Financial Times* (2020).

10.1126/science.abb8952

NEW! Open-Design Upright Microscope

- NEW: Optional motorized or fixed XY stage, or motorized translator
- Open-design microscope with motorized focus
- Quickly configurable based on experimental needs
- Optimized *In vivo* and *In vitro* experimentation on one setup
- Uses standard Olympus objectives
- OCC or DIC transmitted light (LED)
- Epi-fluorescent imaging



BOB™

The Sutter BOB – designed to eliminate the conventional microscope frame – is a versatile, open-design upright microscope platform ideal for slice electrophysiology, widefield fluorescent imaging, two-photon imaging, photostimulation and new techniques just being developed!

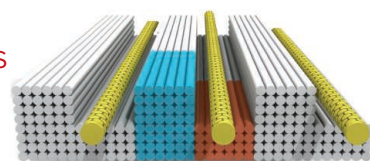
SUTTER INSTRUMENT

PHONE: +1.415.883.0128
FAX: +1.415.883.0572
EMAIL: INFO@SUTTER.COM
WWW.SUTTER.COM

RESEARCH

DNA frame for nanotube devices

Sun et al., p. 874



IN SCIENCE JOURNALS

Edited by Michael Funk

TROPICAL FORESTS

Thermal sensitivity of tropical trees

A key uncertainty in climate change models is the thermal sensitivity of tropical forests and how this value might influence carbon fluxes. Sullivan *et al.* measured carbon stocks and fluxes in permanent forest plots distributed globally. This synthesis of plot networks across climatic and biogeographic gradients shows that forest thermal sensitivity is dominated by high daytime temperatures. This extreme condition depresses growth rates and shortens the time that carbon resides in the ecosystem by killing trees under hot, dry conditions. The effect of temperature is worse above 32°C, and a greater magnitude of climate change thus risks greater loss of tropical forest carbon stocks. Nevertheless, forest carbon stocks are likely to remain higher under moderate climate change if they are protected from direct impacts such as clearance, logging, or fires. —AMS

Science, this issue p. 869

Throughout the tropics, carbon stocks in forests, such as this one in Liberia, will be reduced in response to higher daytime temperatures.

BIOFILMS

Bacteria maintain motile reserves

During biofilm formation, bacterial cells switch from a motile to a generally sessile, matrix-producing state. However, biofilms formed by *Bacillus subtilis* can spread to overtake and kill neighboring colonies of competitor species. Steinberg *et al.* found that a motile subpopulation of cells within *B. subtilis* biofilms was required for the biofilms to spread over foreign objects. This process

required the matrix protein TasA, which stimulated a subset of cells within the biofilm to revert from a matrix-producing to a motile state, thus ensuring that the colony could spread. —AMV

Sci. Signal. **13**, eaaw8905 (2020).

DEVICE TECHNOLOGY

Aligning dense carbon nanotube arrays

Although semiconducting carbon nanotubes (CNTs) are promising candidates to

replace silicon in transistors at extremely small dimensions, their purity, density, and alignment must be improved. Liu *et al.* combined a multiple dispersion sorting process, which improves purity, and a dimension-limited self-alignment process to produce well-aligned CNT arrays on a 10-centimeter silicon wafer. The density is sufficiently high (100 to 200 CNTs per micrometer) that large-scale integrated circuits could be fabricated. With ionic liquid gating, the performance

metrics exceeded those of conventional silicon transistors with similar dimensions. —PDS

Science, this issue p. 850

CELL BIOLOGY

Supramolecular attack particles

Cytotoxic T cells (CTLs) are at the front lines against cancer and chronic infection. T cells kill by secreting caspase-activating granzymes and the pore-forming protein perforin from dense core granules.

However, the structural basis of lethal hit delivery has remained unknown. Balint *et al.* enriched the synaptic output of CTLs to investigate the released form of perforin and granzyme B. They found that CTLs released perforin and granzymes in stable particles called supramolecular attack complexes or SMAPs. The SMAPs were composed of a core shell structure and were assembled in the CTL dense secretory granules before release. The released SMAPs showed an innate ability to kill target cells. —SMH

Science, this issue p. 897

STRUCTURAL BIOLOGY

Transport dependent on context

Transporter proteins move substrates across a membrane, often coupling this activity to cellular ion concentration gradients. For neurotransmitter transporters, which reside in synaptic vesicles that fuse with the plasma membrane after an action potential, transport activity needs to be regulated so that they do not pump out neurotransmitters after vesicle fusion. Using cryo-electron microscopy, Li *et al.* determined the structure of a vesicular glutamate transporter from rat that unveils some of the distinctive features that enable it to function properly in two distinct cellular environments. An allosteric pH sensor, proposed to be a glutamate residue, gates binding of the substrate glutamate and simultaneously permits binding and counterflow of chloride ions. This molecular traffic light allows for a single ion channel to behave appropriately in different contexts. —MAF

Science, this issue p. 893

T CELL MEMORY

Stepping down resident memory lane

The antigen-specific CD8⁺ T cell response to microbial infection includes the

differentiation of a subset of CD8⁺ T cells into tissue-resident memory (T_{RM}) cells that stop circulating and become confined within a nonlymphoid tissue. Kurd *et al.* used single-cell RNA sequencing of mouse CD8⁺ T cells at multiple time points during the first 90 days after viral infection to characterize how this differentiation process unfolds in the small intestine and to track the emergence of heterogeneity among T_{RM} cells. They found evidence for T_{RM} cell precursors in the intestine by 4 days after infection and identified several putative regulators of T_{RM} cell differentiation. The results of this study provide a valuable transcriptomic atlas that will facilitate further investigation into the immune functions provided by T_{RM} cells. —AB

Sci. Immunol. **5**, eaaz6894 (2020).

CORONAVIRUS

What happens next?

Four months into the severe acute respiratory syndrome—coronavirus 2 (SARS-CoV-2) outbreak, we still do not know enough about postrecovery immune protection and environmental and seasonal influences on transmission to predict transmission dynamics accurately. However, we do know that humans are seasonally afflicted by other, less severe coronaviruses. Kissler *et al.* used existing data to build a deterministic model of multiyear interactions between existing coronaviruses, with a focus on the United States, and used this to project the potential epidemic dynamics and pressures on critical care capacity over the next 5 years. The long-term dynamics of SARS-CoV-2 strongly depends on immune responses and immune cross-reactions between the coronaviruses, as well as the timing of introduction of the new virus into a population. One scenario is that a resurgence in SARS-CoV-2 could occur as far into the future as 2025. —CA

Science, this issue p. 860

IN OTHER JOURNALS

Edited by **Caroline Ash**
and **Jesse Smith**

GLACIERS

Continuity in a gap year

The Gravitational Recovery and Climate Experiment (GRACE) and GRACE Follow-On (GRACE-FO) satellite missions have provided measurements of how much mass has been lost from the glaciers of the world since GRACE began operation in 2002. However, GRACE ended in 2017 and GRACE-FO was not launched until 2018, creating a data gap of 1.5 years. Is there an offset of their records? Ciraci *et al.* report that there is not. Using an independent dataset, they show that there is continuity in the satellite mass balance record. This is important because these glaciers, even though they contain far less mass than the Greenland or Antarctic ice sheets, were the biggest contributor to sea-level rise over the 20th century. —HJS

Geophys. Res. Lett. **47**, e2019GL086926 (2020).

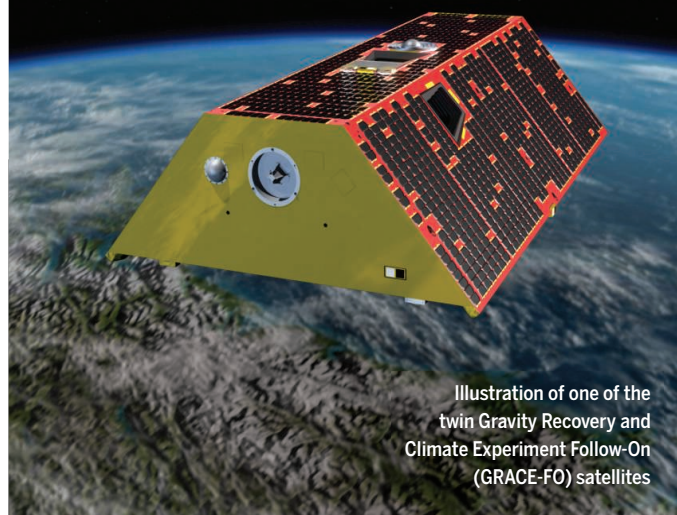


Illustration of one of the twin Gravity Recovery and Climate Experiment Follow-On (GRACE-FO) satellites

IMMUNOGENETICS

HLA genetics and COVID-19

Human leukocyte antigens (HLAs) are proteins encoded by a diverse set of human genes in the major histocompatibility complex. Most people carry between three and six different HLA alleles that show geographically specific distributions. These proteins are important for how the immune system recognizes and mounts immune defenses against infection. Nguyen *et al.* examined how HLA variation affects the cellular immune response to peptides

from human-infecting coronaviruses. The authors found that HLA-B*46:01 had the fewest predicted binding sites to the severe acute respiratory syndrome—coronavirus 2 (SARS-CoV-2) virus, and HLA-B*15:03 showed the greatest capacity to present highly conserved shared SARS-CoV-2 peptides to immune cells. HLA typing may therefore offer valuable information on how coronavirus disease 2019 (COVID-19) might manifest in an individual and help to prioritize treatment options. —LMZ

J. Virol. **10.1128/JVI.00510-20** (2020).



SOIL MICROBIOME

Rehydrating peatland microbiomes

Peatlands are good at storing carbon but are destabilized when they dry out. This happens because oxygen penetrates the dry peat to promote aerobic degradation, releasing carbon and nutrients. Destruction of the mossy overlay follows, which in an intact functioning bog supplies a continuous accumulation of organic matter. Rewetting might restore dried-out peatlands, but whether functioning bog communities that draw down carbon are restored depends on whether microbial communities are restored. Emsens *et al.* surveyed the microbial communities of undrained, dry, and rewetted peatlands. Microbial communities differed between undrained and dry peatland. Rewetted peatlands showed recovery of microbial communities, although loss of organic matter during extended dry times seemed to limit microbiome recovery. —PJH
ISME J. 10.1038/s41396-020-0639-x (2020).

After short periods of drought, rehydration of peat bogs allows recovery of their microbial communities.

CANCER IMMUNOLOGY

Vaccine potential of dendritic cells

The ability of dendritic cells (DCs) to process and present antigens to the immune system makes them intriguing candidates for the development of cancer immunotherapies. Zhou *et al.* investigated whether the CD103⁺ tissue-resident conventional DC1 (cDC1) subset could control tumors in mouse models. DCs were grown and activated in cell culture using an immunostimulant called polyinosinic:polycytidylic acid (poly I:C). The cells were then loaded with tumor antigens and injected into mice to assess their vaccine potential. The authors found that in vitro-generated CD103⁺ cDC1s restrained melanoma and osteosarcoma tumor growth. Tumors could be further hindered if the in vitro-generated cDC1s were combined with immune checkpoint blockade treatment. —PNK

J. Immunother. Cancer **8**, e000474 (2020).

HIGH PRESSURE

Diffuse density determination

Liquid iron at extreme pressure and temperature powers Earth's magnetic field, but its properties are not well constrained. Kuwayama *et al.* used diffuse x-ray scattering to determine the density of molten iron at conditions near those of Earth's outer core. The authors developed a new data analysis method to obtain these values, which can be combined with shock-wave observations to determine the thermal equation of liquid state for iron. The results help to constrain the amounts of alloying elements in Earth's core and provide a new strategy for measuring liquid density at high pressures. —BG

Phys. Rev. Lett. **124**, 165701 (2020).

OPTOMECHANICS

The coolest of vibrations

Shifts in the resonance frequency of micrometer-sized mechanical resonators provide a platform for high-resolution sensing and detection

applications. To exploit optomechanical resonators for quantum technologies, the temperature of the system needs to be very low, ideally well below an occupancy of just one phonon, so that the system is in its ground state. From there, selection and control of the transitions between quantized energy levels provide for quantum-enhanced precision measurement. Qiu *et al.* demonstrate laser cooling of a crystalline silicon optomechanical resonator to its zero-point energy. With a mean thermal phonon occupancy of just ~0.09 quantum, corresponding to ~92% ground state probability and ~7.4 decibels of the zero-point energy, the stage is now set for exquisite quantum-sensing applications as well as fundamental tests of quantum mechanics. —ISO

Phys. Rev. Lett. **124**, 173601 (2020).

DEVELOPMENTAL BIOLOGY

Atlas of gastrulation

We marvel at the progression of a single cell into a complex organism, and microscopy has

long been used to visualize early developmental events. More recently, molecular networks and cellular lineages have been constructed by measuring cells' gene expression profile through single-cell RNA sequencing (scRNA-seq). Sladitschek *et al.* combined scRNA-seq and light-sheet imaging of the ascidian *Phallusia mammillata* to reveal gene expression in all cells of the embryo up to the formation of the multilayered gastrula. Every cell was followed for all cell divisions in 18 cell lineages. The result is a map of cell spatial position and lineage history that reveals embryonic bilateral symmetry and interembryonic variability, as well as the patterned expression of cell adhesion molecules. This resource can be used to examine any number of molecular mechanisms and developmental events occurring in the gastrulating embryo and can serve as model for the spatiotemporal map. —BAP

Cell **10.1016/j.cell.2020.03.055** (2020).

ALSO IN *SCIENCE* JOURNALS

Edited by Michael Funk

;SOCIAL ENVIRONMENT
Social animals need connection

Much research over the past decade or so has revealed that health and lifespan in humans, highly social animals, are reduced with social adversity. We humans are not the only animals that are social, however, and similar research has shown that other social mammals are similarly influenced by isolation and adversity. Snyder-Mackler *et al.* reviewed the relationships between social environment and many aspects of health and well-being across nonhuman mammals and investigated the similarities between these and patterns in humans. They found many of the same threats and responses across social mammals. —SNV

Science, this issue p. 843**MEDICINE**
Correcting blindness

Retinal diseases are a major cause of blindness but numerous developments offer hope that blindness can be reversed. In a Perspective, Dowling discusses advances in visual prostheses to detect light and transmit it to the brain, in using gene therapy to correct inherited blindness, and in applying regenerative approaches, including the implantation of stem cells, to restore visual circuitry. Many of these approaches are being tested in the clinical setting, and if they are successful, sight could be restored in a substantial number of patients blinded by retinal diseases. —GKA

Science, this issue p. 827**CORONAVIRUS**
Drug repurposing

Given the urgent need to find treatments for cases of severe coronavirus disease 2019 (COVID-19), repurposing of existing drugs known to be safe in humans could be a promising

option. In a Perspective, Guy *et al.* discuss the underlying rationale for, and how to prioritize, testing of drugs developed for other applications. Considering the coronavirus replication cycle, certain drugs such as hydroxychloroquine, chloroquine, and remdesivir could have promising activity against COVID-19. However, the authors urge caution in relying on studies in cells and results from preliminary clinical trials involving small numbers of patients. Moreover, the side effects of some of these drugs should also be balanced against possible gains. —GKA

Science, this issue p. 829**PLANT SCIENCE**
Fungal disease meets its match

Fusarium head blight (FHB), caused by a fungus, reduces wheat crop yield and introduces toxins into the harvest. From the assembly of the genome of *Thinopyrum elongatum*, a wild relative of wheat used in breeding programs to improve cultivated wheat, Wang *et al.* cloned a gene that can address both problems (see the Perspective by Wulff and Jones). The encoded glutathione S-transferase detoxifies the trichothecene toxin and, when expressed in wheat, confers resistance to FHB. —PJH

Science, this issue p. 844;
see also p. 822**WATER RESOURCES**
Dowsing for danger

Arsenic is a metabolic poison that is present in minute quantities in most rock materials and, under certain natural conditions, can accumulate in aquifers and cause adverse health effects. Podgorski and Berg used measurements of arsenic in groundwater from ~80 previous studies to train a machine-learning model with globally continuous predictor variables,

including climate, soil, and topography (see the Perspective by Zheng). The output global map reveals the potential for hazard from arsenic contamination in groundwater, even in many places where there are sparse or no reported measurements. The highest-risk regions include areas of southern and central Asia and South America. Understanding arsenic hazard is especially essential in areas facing current or future water insecurity. —MAF

Science, this issue p. 845;
see also p. 818**TOPOLOGICAL OPTICS**
Topological insulators go nonlinear

Whereas solid-state insulators tend to be fixed by material properties, photonic topological insulators can be designed at will to mimic a variety of scenarios and complex interactions. Mukherjee and Rechtsman go beyond the linear optical regime that has been studied to date and show that photonic topological insulators can also exhibit nonlinear optical features (see the Perspective by Ablowitz and Cole). Their array of laser-written waveguides can support solitons, which are also found to exhibit topological features, performing cyclotron-like orbits associated with the topology of the lattice. The nonlinear properties provide a rich playground for further exploration, with the possibility of mimicking other interacting bosonic systems. —ISO

Science, this issue p. 856;
see also p. 821**METROLOGY**
Good timing for microwave technology

Timing standards around the world define the second using atomic clocks, specifically the microwave frequencies emitted

from trapped atoms. Optical clocks, which are based on optical transitions of atoms, operate at much higher frequency and have been shown to exhibit better stability. Nakamura *et al.* demonstrate a framework that carries the improved stability of the optical domain over to microwaves (see the Perspective by Curtis). In addition to contributing to the eventual redefinition of the second based on optical clocks, this work will also lead to improvements in microwave-based technologies such as astronomical imaging and geodesy through very long baseline interferometry, radar, communications, and navigation systems. —ISO

Science, this issue p. 889;
see also p. 825**DEVICE TECHNOLOGY**
DNA bricks build nanotube transistors

Semiconducting carbon nanotubes (CNTs) are an attractive platform for field-effect transistors (FETs) because they potentially can outperform silicon as dimensions shrink. Challenges to achieving superior performance include creating highly aligned and dense arrays of nanotubes as well as removing coatings that increase contact resistance. Sun *et al.* aligned CNTs by wrapping them with single-stranded DNA handles and binding them into DNA origami bricks that formed an array of channels with precise intertube pitches as small as 10.4 nanometers. Zhao *et al.* then constructed single and multichannel FETs by attaching the arrays to a polymer-templated silicon wafer. After adding metal contacts across the CNTs to fix them to the substrate, they washed away all of the DNA and then deposited electrodes and gate dielectrics. The FETs showed high on-state performance and fast on-off switching. —PDS

Science, this issue p. 874, p. 878

CHEMICAL PHYSICS

Electronic and nuclear dynamics in one

Because of the complex, ultrafast interplay between nuclear and electronic degrees of freedom, probing both nuclear and electronic dynamics in excited electronic states within a single time-resolved experiment is a great challenge. Yang *et al.* used ultrafast electron diffraction in combination with *ab initio* non-adiabatic molecular dynamics and diffraction simulations to study the relaxation dynamics of isolated pyridine molecules after photoexcitation to the S_1 state (see the Perspective by Domcke and Sobolewski). They showed that electronic state evolution and molecular structural changes can be recorded simultaneously and independently by tracing a transient signal in small-angle inelastic scattering and large-angle elastic diffraction, respectively. —YS

Science, this issue p. 885;
see also p. 820

POLLINATORS

Bumble bee gardeners

Bumble bees rely heavily on pollen resources for essential nutrients as they build their summer colonies. Therefore, we might expect that annual differences in the availability of these resources must simply be tolerated, but Pashalidou *et al.* made observations suggesting that bees may have strategies to cope with irregular seasonal flowering (see the Perspective by Chittka). When faced with a shortage of pollen, bumble bees actively damaged plant leaves in a characteristic way, and this behavior resulted in earlier flowering by as much as 30 days. Experimenters were not able to fully replicate the results with their own damage, suggesting that there is a distinct method that the bees use to stimulate earlier flowering. —SNV

Science, this issue p. 881;
see also p. 824

BONE

Screening for side-effect susceptibility

Osteoporosis is typically treated with nitrogen-containing bisphosphonates (N-BPs) to inhibit bone resorption; however, N-BPs can cause serious side effects, including osteonecrosis and fracture. To understand susceptibility to these side effects, Surface *et al.* investigated the role of the gene *ATRAID* in mediating response to N-BPs. Mice deficient in this gene had weaker bone and did not respond to N-BP treatment in models of osteoporosis. Patients with coding variants in *ATRAID* taking N-BPs presented with fractures and osteonecrosis of the jaw. The authors determined that *ATRAID* is necessary for inhibition of osteoclast function by N-BPs. It may thus be prudent to screen patients for variants in *ATRAID* to avoid N-BP side effects. —CC

Sci. Transl. Med. **12**, eaav9166 (2020).

REVIEW SUMMARY

SOCIAL ENVIRONMENT

Social determinants of health and survival in humans and other animals

Noah Snyder-Mackler, Joseph Robert Burger, Lauren Gaydosh, Daniel W. Belsky, Grace A. Noppert, Fernando A. Campos, Alessandro Bartolomucci, Yang Claire Yang, Allison E. Aiello, Angela O'Rand, Kathleen Mullan Harris, Carol A. Shively, Susan C. Alberts, Jenny Tung*

BACKGROUND: The social environment shapes human health, producing strong relationships between social factors, disease risk, and survival. The strength of these links has drawn attention from researchers in both the social and natural sciences, who share common interests in the biological processes that link the social environment to disease outcomes and mortality risk. Social scientists are motivated by an interest in contributing to policy that improves human health. Evolutionary biologists are interested in the origins of sociality and the determinants of Darwinian fitness. These research agendas have now converged to demonstrate strong parallels between the consequences of social adversity in human populations and in other social mammals, at least for the social processes that are most analogous between species. At the same time, recent studies in experimental

animal models confirm that socially induced stress is, by itself, sufficient to negatively affect health and shorten life span. These findings suggest that some aspects of the social determinants of health—especially those that can be modeled through studies of direct social interaction in nonhuman animals—have deep evolutionary roots. They also present new opportunities for studying the emergence of social disparities in health and mortality risk.

ADVANCES: The relationship between the social environment and mortality risk has been known in humans for some time, but studies in other social mammals have only recently been able to test for the same general phenomenon. These studies reveal that measures of social integration, social support, and, to a lesser extent, social status independently

predict life span in at least four different mammalian orders. Despite key differences in the factors that structure the social environment in humans and other animals, the effect sizes that relate social status and social integration to natural life span in other mammals align with those estimated for social environmental effects in humans. Also like humans, multiple distinct measures of social integration have predictive value, and in the taxa examined thus far, social adversity in early life is particularly tightly linked to later-life survival.

Animal models have also been key to advancing our understanding of the causal links between social processes and health. Studies in laboratory animals indicate that socially induced stress has direct effects on immune function, disease susceptibility, and life span. Animal models have revealed pervasive changes in the response to social adversity that are detectable at the molecular level. Recent work in mice has also shown that socially induced stress shortens natural life spans owing to multiple causes, including atherosclerosis. This result echoes those in humans, in which social adversity predicts increased mortality risk from almost all major causes of death.

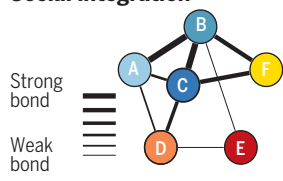
OUTLOOK: Although not all facets of the social determinants of health in humans can be effectively modeled in other social mammals, the strong evidence that some of these determinants are shared argues that comparative studies should play a frontline role in the effort to understand them. Expanding the set of species studied in nature, as well as the range of human populations in which the social environment is well characterized, should be a priority. Such studies have high potential to shed light on the pathways that connect social experience to life course outcomes as well as the evolutionary logic that accounts for these effects. Studies that draw on the power and tools afforded by laboratory model organisms are also crucial because of their potential for identifying causal links. Important research directions include understanding the predictors of interindividual and intersocietal differences in response to social adversity, testing the efficacy of potential interventions, and extending research on the physiological signatures of social gradients to the brain and other tissues. Path-breaking studies in this area will not only integrate results from different disciplines but also involve cross-disciplinary efforts that begin at study conception and design. ■

The list of author affiliations is available in the full article online.

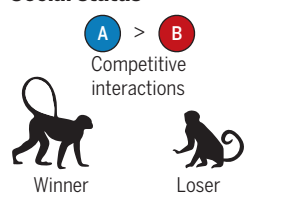
*Corresponding author. Email: jenny.tung@duke.edu

Cite this article as N. Snyder-Mackler et al., *Science* 368, eaax9553 (2020). DOI: 10.1126/science.aax9553

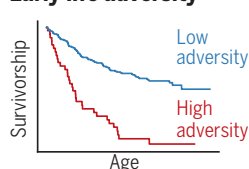
Social integration



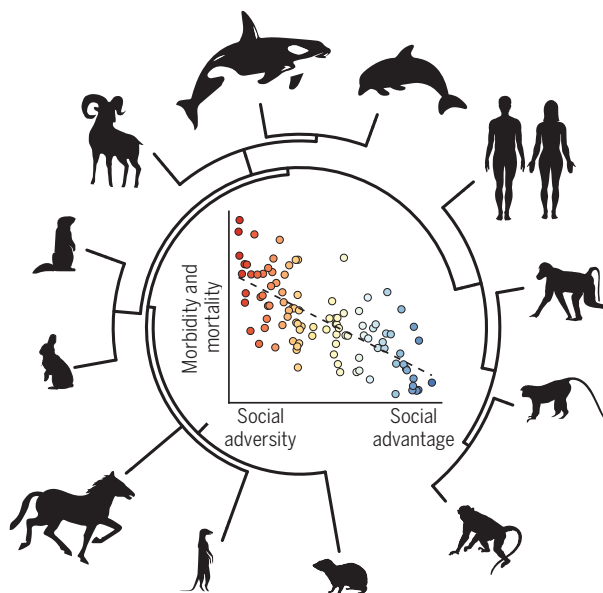
Social status



Early life adversity



A comparative perspective on the social determinants of health. Social adversity is closely linked to health and mortality outcomes in humans, across the life course. These observations have recently been extended to other social mammals, in which social integration, social status, and early-life adversity have been shown to predict natural life spans in wild populations and molecular, physiological, and disease outcomes in experimental animal models.



REVIEW

SOCIAL ENVIRONMENT

Social determinants of health and survival in humans and other animals

Noah Snyder-Mackler^{1,2,3,4}, Joseph Robert Burger^{1,5,6,7}, Lauren Gaydos^{1,8}, Daniel W. Belsky^{1,5,9,10}, Grace A. Noppert^{1,5,11,12,13}, Fernando A. Campos^{1,14,15}, Alessandro Bartolomucci¹⁶, Yang Claire Yang^{1,12,17,18}, Allison E. Aiello^{1,12,19}, Angela O'Rand^{1,5,11}, Kathleen Mullan Harris^{1,12,17}, Carol A. Shively^{1,20}, Susan C. Alberts^{1,2,5,11,14,21}, Jenny Tung^{1,2,5,11,14,21*}

The social environment, both in early life and adulthood, is one of the strongest predictors of morbidity and mortality risk in humans. Evidence from long-term studies of other social mammals indicates that this relationship is similar across many species. In addition, experimental studies show that social interactions can causally alter animal physiology, disease risk, and life span itself. These findings highlight the importance of the social environment to health and mortality as well as Darwinian fitness—outcomes of interest to social scientists and biologists alike. They thus emphasize the utility of cross-species analysis for understanding the predictors of, and mechanisms underlying, social gradients in health.

In social mammals, including our own species, social conditions powerfully shape the environment that individuals experience from day to day. Adverse social experiences, in particular, elicit biological responses across social species that influence health and aging across the life span (1). It is therefore unsurprising that dimensions of the social environment—particularly measures of socioeconomic status, social integration, and early-life adversity—are among the strongest and most consistent predictors of health and survival outcomes (Fig. 1). For example, differences in socioeconomic status in the United States (as measured by income) can translate

to differences of a decade or more of life expectancy (2), and low occupational status translates to ~2 years of reduced life span across seven high-income countries (3). Similarly, low social integration predicts a ~50% increase in all-cause mortality risk in humans, an effect that rivals or exceeds mortality risk associated with obesity, alcoholism, moderate smoking, or sedentary living (4).

These observations raise a natural question: What are the biological processes that account for the strong association between the social environment, disease, and mortality risk? This question is relevant to improving disease prediction, prevention, and targeting interventions; understanding the causes and consequences of social inequality; and investigating the evolution of social group living and its relevance to health. It is also timely. In the past two decades, socioeconomic disparities in mortality have become steeper in the United States (5, 6). Aging populations have also highlighted the negative effects of social isolation in the elderly (7, 8); in response, the United Kingdom appointed its first Minister of Loneliness in 2018, and the World Health Organization has launched initiatives to focus attention on the social determinants of health. Prospective studies have placed early-life conditions at the root of some of these observations (9, 10). The increasing concern about social disparities in health indicates that the current array of measures being used to study and mitigate social gradients are incomplete. Understanding the biology underlying social environmental effects on health—especially physiological changes that precede disease itself—promises to provide new opportunities for effective intervention.

Addressing this question has been challenging for at least two reasons. First, considerable evidence, drawn almost entirely from animal

models, supports the hypothesis that social interactions directly affect health outcomes (the “social causation” hypothesis) (11, 12). However, social gradients in human health can also be explained by other environmental mediators (such as diet, smoking, and health care access) (13–15), and in some cases, poor health can cause individuals to experience more adverse social exposures (“health selection”). In many studies of humans, including some that have been foundational for characterizing the effects of social adversity, considerable uncertainty surrounds the relative contribution of social causation versus health selection (14, 16–19). This challenge arises because experimental studies of exposure to many sources of social adversity are nearly impossible in humans. The problem is further compounded by the absence of information about social and biological conditions before the start of many key studies and by the interdependence between social gradients and health over time. Longitudinal datasets that include baseline measures partially address these challenges (16, 20, 21) but still cannot unambiguously disentangle causal pathways because of the difficulty of excluding the effects of correlated or confounding variables (such as time-varying confounders) (6, 22). However, some quasi-experimental studies have found that modest increases in measures of socioeconomic status (income and/or neighborhood conditions) can positively affect physical and mental health, emphasizing the need for further study (23–25).

Second, associations between the social environment and health pose a challenge to typical strategies for studying the biological mechanisms of disease. Social adversity is linked to a remarkably broad set of conditions, including diseases as distinct as tuberculosis, diabetes, cardiovascular disease, and cancer (Fig. 1, D to F). The fact that so many different physiological systems are socially patterned makes choosing an appropriate animal, tissue, or cellular model difficult. This problem is further complicated by the fact that studies of the social environment minimally require social interaction in groups or communities, meaning that social cues cannot be readily modeled in individually housed organisms or cell lines. Even assuming a social causation model, the health consequences of social adversity fit poorly into classical host-agent-environment models, which represent the typical biological approach to studying disease causation (26, 27). Studies have instead tended to discuss the social environment in terms of a general “predisposing risk factor,” “social exposure,” or source of “accumulated wear and tear” (28–30). These are useful conceptual models but provide little guidance for traditional studies of biological mechanisms.

Thus, despite broad interest in the biological correlates and consequences of social adversity, the mechanisms, processes, and pathways

¹Social and Biological Determinants of Health Working Group, NC, USA. ²Department of Evolutionary Anthropology, Duke University, Durham, NC, USA. ³Department of Psychology, University of Washington, Seattle, WA, USA. ⁴Center for Evolution and Medicine, Arizona State University, Tempe, AZ, USA. ⁵Population Research Institute, Duke University, Durham, NC, USA. ⁶Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, AZ, USA. ⁷Institute of the Environment, University of Arizona, Tucson, AZ, USA. ⁸Center for Medicine, Health, and Society, Vanderbilt University, Nashville, TN, USA. ⁹Department of Epidemiology, Columbia University Mailman School of Public Health, New York, NY, USA. ¹⁰Robert N. Butler Columbia Aging Center, Columbia University Mailman School of Public Health, New York, NY, USA. ¹¹Center for Population Health and Aging, Duke University, Durham, NC, USA. ¹²Carolina Population Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. ¹³Center for the Study of Aging and Human Development, Duke University, Durham, NC, USA. ¹⁴Department of Biology, Duke University, Durham, NC, USA. ¹⁵Department of Anthropology, University of Texas at San Antonio, San Antonio, TX, USA. ¹⁶Department of Integrative Biology and Physiology, University of Minnesota, Minneapolis, MN, USA. ¹⁷Department of Sociology, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. ¹⁸Lineberger Cancer Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. ¹⁹Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. ²⁰Comparative Medicine Section, Department of Pathology, Wake Forest School of Medicine, Winston-Salem, NC, USA. ²¹Institute of Primate Research, Nairobi, Kenya. *Corresponding author. Email: jenny.tung@duke.edu

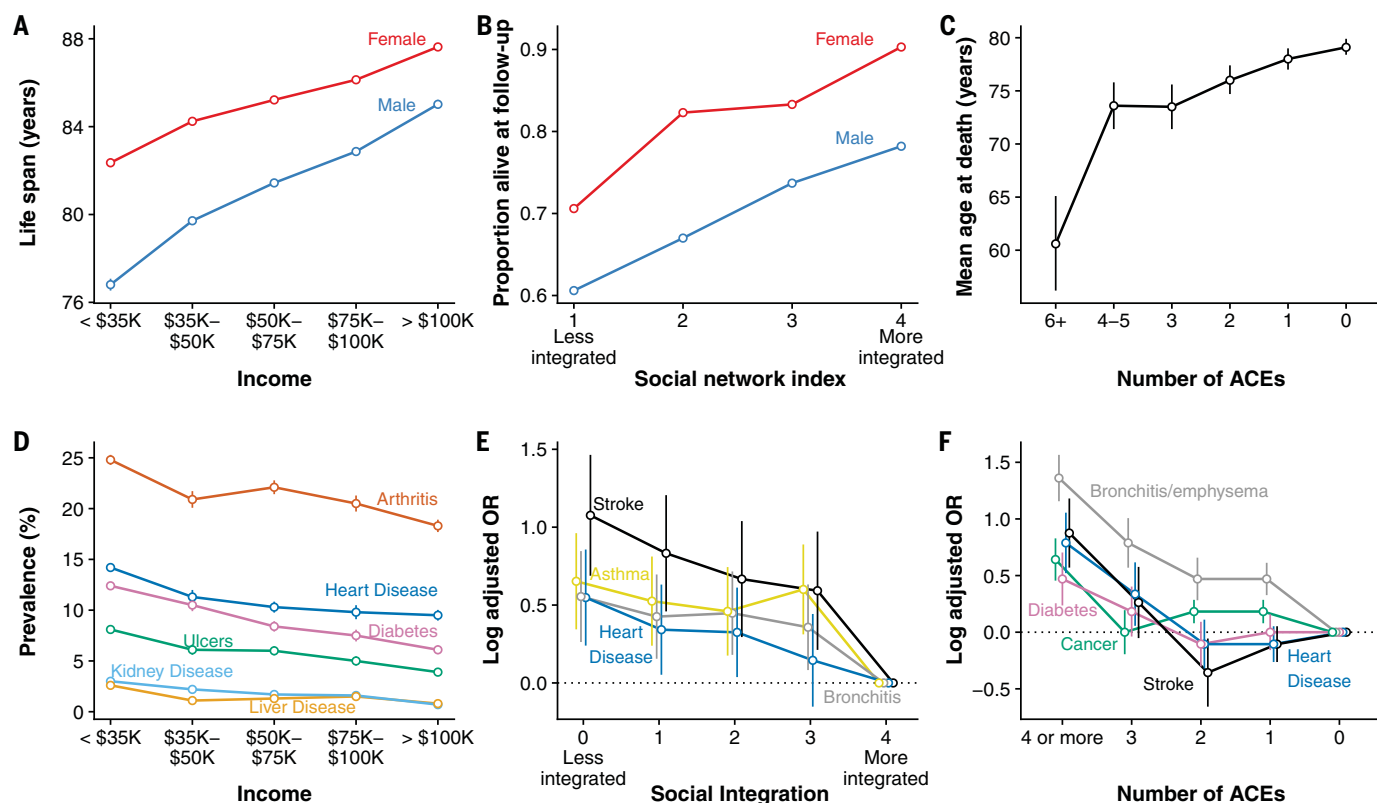


Fig. 1. Social adversity predicts morbidity and mortality in humans. (A to F) The largest data sets on the health correlates of social adversity come from human populations. Together, they demonstrate that high social adversity is a major predictor of [(A) to (C)] life expectancy and [(D) to (F)] susceptibility to a broad range of diseases. (A) Expected life span at age 40 for men and women in the United States as a function of income at age 40 ($n = 1.4$ billion person-years) (2). (B) Proportion of study subjects alive after a 9-year follow up, for adult men and women in Alameda County, California, as a function of a composite index of social relationships ($n = 6298$ individuals) (46). (C) Mean age at death as a function of early adversity in the ACEs study on adult patients at the Kaiser Permanente

San Diego Health Appraisal Clinic ($n = 17,337$ individuals, $n = 1539$ who had died by follow up) (173). (D) Disease prevalence among adult Americans by income based on the 2015 Centers for Disease Control National Health Interview Survey ($n = 242,501$ individuals) (174). (E) Disease risk (log odds ratios adjusted for age, sex, and race) as a function of a composite measure of social integration for adult men and women in the United States in the National Health and Nutrition Examination Survey III ($n = 18,716$ individuals) (31). (F) Disease risk (log odds ratios adjusted for age, sex, race, and educational attainment) by number of ACEs for patients visiting Kaiser Permanente's San Diego Health Appraisal Clinic ($n = 9508$ individuals) (9).

through which they arise have remained unclear. However, new evidence has been crucial for moving toward clarity on these questions. First, research in other social mammals indicate that social gradients in human health are part of a long evolutionary legacy of social living, at least at the level of local social interactions among coresident individuals (Figs. 2 and 3 and Box 1). These findings suggest that the consequences of social adversity transcend the effects of the modern human environment and point to evolutionary comparative studies as a source of important insight. Second, emerging data sets, especially controlled experimental studies in other social mammals, strongly support direct social environmental effects on physiological function (social causation). Together with the release of unprecedentedly large, integrated data sets from human populations (2, 3, 31), these findings lay the groundwork for understanding how social adversity makes us vulnerable.

Here, we review key themes emerging from this evidence, with an emphasis on recent work that highlights the role of social experience across the life course and findings of shared interest across disciplines. Because this intersection necessarily links to multiple fields, we do not attempt to summarize the full scope of research on either the social determinants of health in humans (which also involve socioeconomic structures not applicable to animal models) or the fitness consequences of social behavior in humans and other animals; instead, we refer readers to excellent reviews, with a within-discipline focus here (6, 11, 32–36). Our goal in this Review is to emphasize emerging parallels and insights from studies of social mammals, in the context of observations initially made in human populations. We focus on social mammals—particularly those that obligately live in groups—because of their close evolutionary relationship to humans. However, social environmental effects on health and fit-

ness have also been of interest in other species, especially birds and social insects (37). The degree to which these more distantly related species can be used to understand the social determinants of health in humans remains an important question for future work.

Social adversity and mortality in social mammals

In the social sciences, research on the social determinants of health is motivated by an interest in contributing to policy that reduces health disparities and improves human health span, life span, or life expectancy. This work has a long tradition; social gradients have been described in the sociological literature for at least 120 years (38). In parallel, evolutionary biologists and behavioral ecologists study social interactions with an eye toward understanding the origins of sociality and its consequences for reproductive fitness. This research program is also old; Darwin himself puzzled over the

adaptive value of social behavior (39), which is thought to have imparted strong enough selective pressure to drive major morphological and physiological innovations, including advanced cognitive abilities in humans and other primates (40, 41).

Over the past decade, the historically distinct agendas of social science and evolutionary biology have begun to converge. In particular, several long-term studies in wild social mammals now contain enough data to support full life course analyses and have revealed unexpectedly strong links between the social environment and mortality risk that parallel those from long-term studies in humans (Figs. 2 and 3). These findings simultaneously connect to the motivating questions for evolutionary biologists—life span is often the most important predictor of Darwinian fitness (reproductive success, the determinant of an individual genome's representation in future generations) in long-lived mammals (42)—and place observations in humans on a biological continuum with other species. Together, they illustrate several patterns that consistently shape social gradients in humans and other social mammals and provide crucial justification for studying the biology of social gradients in other species.

Below, we review the evidence for this convergence in connection with three dimensions of the social environment: (i) social integration, defined as an individual's ability to invest in and maintain affiliative or supportive interactions (whether shaped by intrinsic ability or by the constraints of its environment) (43); (ii) social status, a construct that captures stable or semistable differences in access to resources, whether material (such as food, health care or access to mates) or otherwise (such as psychological capital or social support); and (iii) early-life adversity, with an emphasis on social and familial adversity that occurs during sensitive periods in development. In animals, all three dimensions are captured through observations of direct social interactions. This is an important point of divergence from human studies, in which researchers often measure engagement in larger social, cultural, and economic structures that can knit individuals into a shared socioeconomic framework, even if they never meet. Such structures do not have a clear parallel in animal models; for example, it is difficult to put animals from different geographic locations on a single status scale because they do not interact (although it is possible to ask whether relatively low-status animals in different groups do worse on average, and some researchers have investigated the relative "status" of distinct social groups in relation to one another) (44). The relative simplicity of nonhuman animal societies is thus both an advantage—it rules out some potential confounders and causal

pathways that complicate interpretation in humans—and a limitation, as not all aspects of the social determinants of health can be effectively modeled in nonhuman animals. Nevertheless, as in humans, multiple, non-mutually exclusive pathways link social factors to each other and to health and fitness outcomes (Box 1).

Social integration and survival

In humans, the evidence for a link between social isolation and mortality risk is extensive and remarkably consistent across geographically, temporally, and socioeconomically diverse populations (although the current data are largely limited to societies in the developed world) (4, 45). The earliest population-based studies to investigate this relationship estimated that social integration increased the odds of survival by 30 to 80% (odds ratio between 1.3 and 1.8) (46). Recent meta-analyses have included several orders of magnitude more study subjects but nevertheless encompass these original values, with odds ratios ranging from 1.19 to 1.91 depending on measurement approach and inclusion criteria (4, 8).

Emerging results from wild mammals are strikingly similar to those in humans. The first wild animal study to demonstrate a relationship between individual-based measures of social integration and adult survival, in wild baboons, was published a decade ago (47). Since then, similar results have been reported for a variety of other social mammals, including independent replication in a second population of baboons (Fig. 2) (48). In some species, juvenile survival may also be linked to the ability to socially integrate into mixed-age social groups (49, 50). An important caveat to these studies is that some are based on very small sample sizes, others do not control for group size or population density (which could affect survival through mechanisms other than the opportunity for affiliative social interactions) (51), and the direction of causation cannot be easily determined. Further, a few exceptions stand out. For example, in yellow-bellied marmots, females who were more well-integrated into a social network died earlier; this difference from other social mammals may be linked to the fact that social group living is not obligate in this species, unlike the others that have been studied (52, 53). In other cases, the results depend on specific measures of social integration: In blue monkeys, females who maintained strong and consistent social bonds with the same partners lived longest, but those with strong and inconsistent bonds fared the worst (54). Thus, caution should be exercised in painting a homogeneous picture across all social mammals. Nonetheless, the pattern of greater survival with greater social integration appears relatively consistent in studies of wild

mammals thus far and is remarkably close to the effect sizes in humans, with odds ratios in the range of 1.23 to 1.72 (Fig. 2). These studies include representatives from five mammalian orders and capture multiple independent evolutionary transitions to social living (in primates, rodents, odd-toed ungulates, even-toed ungulates, and hyracoidea or their ancestors) (55). These observations suggest a convergent relationship between affiliative social interactions and survival that is detectable across tens of millions of years of evolutionary time.

In keeping with studies in humans, this pattern is evident despite substantial variation in measurement approaches. Although all measures are based on direct observation of social interactions, some have relied on social network analyses of affiliative interactions or proximity to "neighbors," whereas others have focused on pairwise interactions (such as bond strength, consistency, or the relative frequency of interactions). These studies represent a mix of what are called, in studies of humans, "structural" measures (such as the number of social ties or the position of an individual in a network) and "functional" measures (such as the extent to which social ties provide particular resources, including perceived social support in humans). In humans, structural and functional measures are only moderately correlated with each other but have similar associations with survival, and multidimensional measures make the best predictors (4, 43, 56, 57). No study has yet examined the relationship between structural and functional measures in wild mammals, although both types of data have been analyzed. For example, the thermoregulatory benefit of social huddling in Barbary macaques (58) and vervet monkeys (59) is a functional measure; network centrality in bighorn sheep (60) and orcas (61) is a structural measure (centrality is a measure of the contribution an individual makes to a social network's overall connectivity) (62). However, several studies indicate that measures of affiliative social relationships vary in predictive power (63, 64), and a recent comparative analysis in rhesus macaques points to the particular importance of bond strength and consistency, as opposed to affiliative behavior (such as grooming) per se (65). As the number and power of available studies grow, comparisons of structural and functional measures across species should further refine the aspects of social integration that are most consistently important.

Social status and survival

Like social integration and support, the overall link between socioeconomic status and survival rates in human populations is well established and cuts across cultural and national boundaries (66, 67). The earliest data on this phenomenon, from the United Kingdom starting in 1931, showed that the risk of death from heart

disease was more than twofold higher for men in the lowest social class than in the highest (68). Fifty years later, the Whitehall studies of British civil servants revealed more than a threefold difference among white-collar British workers (69). Today, we know that low socioeconomic status is linked to increased mortality risk from nearly all causes, including chronic disease and infectious disease as well as accidents and violent death (Fig. 1) (2, 34, 66, 70, 71). The consistency of this relationship over time and space has motivated some researchers to label socioeconomic status inequalities as a “fundamental cause” of disease (28).

Social status in other social mammals is much simpler. Hierarchies do not extend beyond the members of a coresident social group or population, and a single measure of status—

typically dominance rank, which is commonly defined as the ability to win social conflicts or to displace conspecifics from resources (1)—is usually sufficient to capture stable differences in resource access (although within species, dominance rank can be sex specific). However, in other social mammals, too, social status is often linked to survival and can predict physiological differences that strongly parallel those observed in humans (32, 33, 72–74). Despite long-standing interest in its causes and consequences, the relationship between social status and fertility has been more intensively studied than its relationship with survival (75–77), and the literature on social status and life span remains somewhat biased toward long-term studies of primates. Nonetheless, results are generally consistent with those observed in humans; to date, studies of wild rab-

bits (78), meerkats (79), baboons (47, 80), rhesus macaques (81), and long-tailed (cynomolgus) macaques (82) all show a survival advantage to higher social rank (although not always in a linear fashion).

As with studies of social integration and survival, comparative analyses may help identify factors that influence the link between social status and survival. For physiological outcomes, comparative studies in animals already emphasize that the costs of low status are moderated by social context. Low-status animals tend to exhibit higher levels of stress-associated glucocorticoid hormones when they belong to strictly enforced hierarchies and lack access to social support (83), suggesting that social status and social integration may have interrelated effects on health outcomes (Box 1). One study of wild female baboons showed

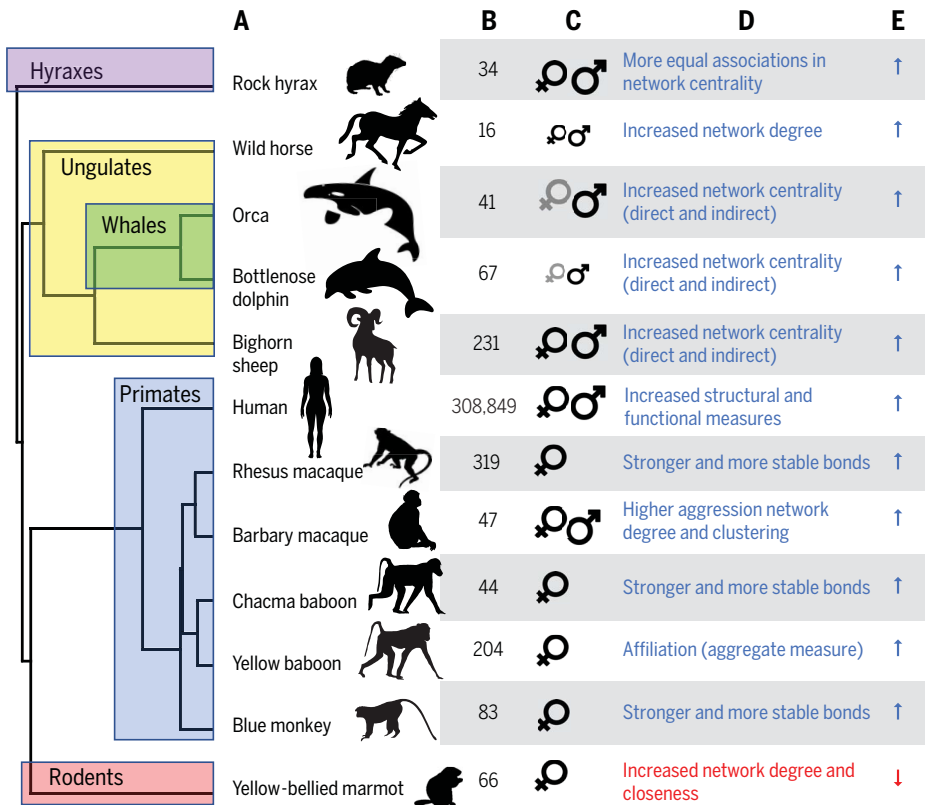


Fig. 2. Social integration and survival in wild social mammals. All cases shown are based on data from natural populations, with the exception of rhesus macaques (65), for which data are from a provisioned free-ranging population. (A) The social integration-survival relationship has been evaluated in at least 12 species, including humans, which together represent multiple independent transitions to social group living (55). The mammal supertree is from (175). (B) Sample sizes and (C) sex studied. Large symbols indicate adults; small symbols indicate juveniles. Sample size for humans is based on a meta-analysis of 148 studies. Where both sexes were investigated, significant results are shown in black and nonsignificant results in gray. (D) Measure of social integration tested. (E) Direction of the observed effect. Blue arrows correspond to improved survival with greater integration and support; red arrow

corresponds to reduced survival with greater integration and support. For Barbary macaques, affiliative networks were unrelated to survival; for orca, social integration predicted survival in males only in limited resource years. We excluded several studies of wild mammals that focused on social group size as the measure of social support and integration [cheetahs (176), wolves (177), voles (178), and bats (179)] because the effects of social factors cannot be disentangled from the effects of other density-dependent factors (such as degree of resource competition and between-group competition). Data are from the following sources: rock hyrax, (180); wild horse, (50); orca, (61); bottlenose dolphin, (49); bighorn sheep, (60); human, (4); rhesus macaque, (65); Barbary macaque, (181); chacma baboon, (47); yellow baboon, (48); blue monkey, (54); yellow-bellied marmot, (53).

that social status did not directly predict survival, but social affiliation did. However, higher-ranking females were more socially affiliated to males, suggesting an indirect effect of social status on survival (48). The survival advantage for dominant meerkats is also explained by the effects of social status on social integration: Subordinates were less well-integrated into the group and hence more exposed to extrinsic mortality risks, such as predators (79). Last, studies in social mammals highlight how variation in the nature of social status attainment and maintenance can produce distinct biological outcomes (84). For example, some hierarchies are determined by physical strength and are therefore dynamic over time (such as male baboons and male red deer), whereas others are largely determined by the social status of close kin (such as female baboons and female spotted hyenas). In the latter case, hierarchies can persist over multiple generations (85, 86), providing what is perhaps the closest non-human analog to structurally embedded social hierarchies in humans.

The long-term effects of early-life adversity
Early development is a period of substantial sensitivity to environmental adversity, including social as well as physical hardship. In humans, extensive evidence supports a relationship between social adversity in early-life and later-life health outcomes, including reproductive timing, cardiovascular disease, viral infection, and premature mortality (87–90). For example, low socioeconomic status in early life is associated with a more than twofold increased probability of early-onset coronary heart disease, even among study subjects who achieved high socioeconomic status as adults (91). Similarly, racial and ethnic minorities who climb the social ladder to higher status in young adulthood nevertheless experience early adversity-associated costs to health (92–95). Such observations suggest that the social roots of later-life health gradients can be established many years earlier and may be refractory to later life change, perhaps because of biological embedding (30).
Early-life effects are also well studied in other animals [including in many nonmam-

mals (96, 97)]. However, although the early-life social environment has long been linked to physiological, growth-associated, and cognitive traits (98), its relationship to adult health and survival—especially after a long intervening period—has only recently been investigated in natural populations. In the first study in animals to use the adverse childhood experiences (ACEs) framework, which tallies the number of discrete insults experienced early in life (an ACE represents a potentially traumatic or developmentally disruptive environmental exposure in early life, such as physical abuse or familial separation in humans), yellow baboon females who experienced more early-life adversity were shown to experience substantially shorter life spans (99). Females who experienced three or more major insults (of six studied, including low social status, maternal social isolation, maternal loss, high resource competition, a short interval until the birth of a younger sibling, and early-life drought) died approximately a decade earlier than those who experienced none, an effect size even larger

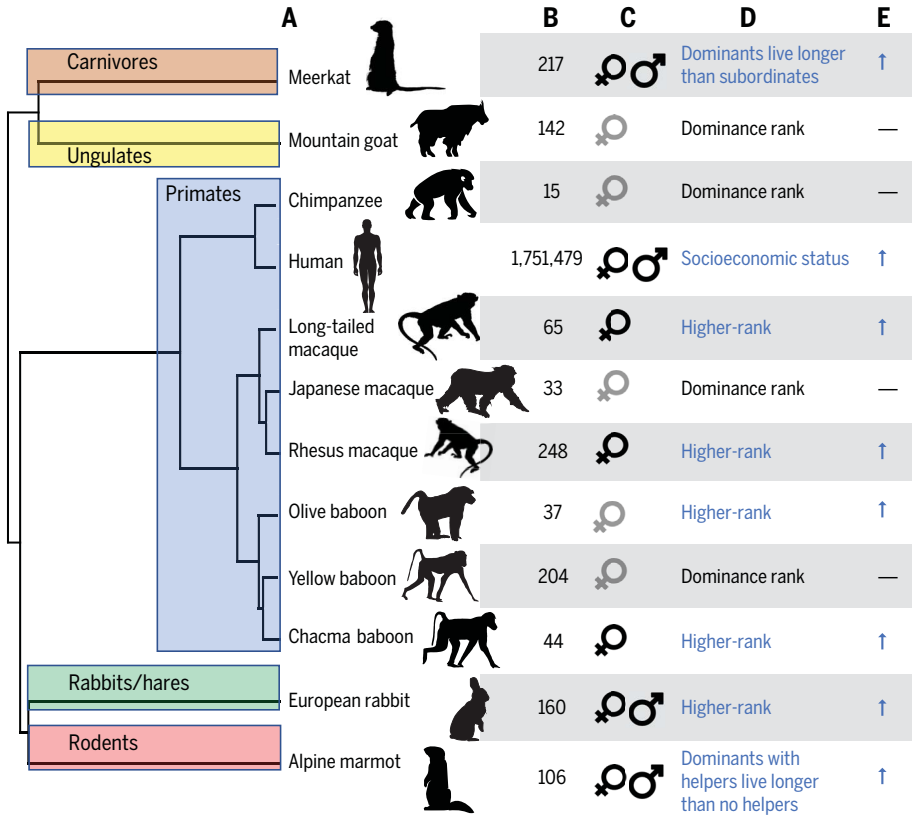


Fig. 3. Social status and survival in wild social mammals. All cases shown are based on data from natural populations. (A) The social status–survival relationship has been evaluated in at least 12 species, including humans, which together represent multiple transitions from solitary to social living (in carnivores, even-toed ungulates, primates, rabbits and hares, and rodents) (55). The mammal supertree is from (175), with modifications based on (182). (B) Sample sizes and (C) sex studied. Sample size for humans is based on a meta-analysis of 48 studies. Where both sexes were investigated, significant results are shown in black and nonsignificant results in gray. (D) Measure of social status tested. (E) Direction of the observed effect. Blue arrows correspond to improved survival with higher social status or rank; dashes correspond to no relationship between survival and social status or rank, as reported based on the authors’ threshold for statistical significance. Data are from the following sources: meerkat, (79); mountain goat, (183); chimpanzee, (184); human, (3); long-tailed macaque, (82); Japanese macaque, (185); rhesus macaque, (81); olive baboon, (186); yellow baboon, (47); chacma baboon, (80); European rabbit, (78); alpine marmot, (187).

than those documented in human populations (Fig. 1F). Most of the sources of early adversity had a social component, and the two with the largest predictive effects—maternal loss and the birth of a close-in-age younger sibling—specifically point to the importance of mothers as a source of early-life social support. Recent work in wild spotted hyenas, a highly social carnivore, corroborates these findings (100). In hyenas, a cumulative adversity index incorporating maternal social status, maternal loss in the infant–juvenile period, and an animal's own deviation from expected social status early in life also strongly predicts life span, again on a time scale of years.

These results both fit with and enrich models of early adversity developed for human populations that attempt to account for ACEs-related results (101). For example, consistent with the accumulation of risks model (102, 103), they indicate that sequential deleterious exposures combine to have especially negative effects. However, although sources of early adversity in humans are often correlated—for example, children living in poverty are also likely to live in households with a missing parent (104)—in wild animal populations, correlations between different sources of adversity may be weak or absent altogether (99). This structure facilitates examination of the cumulative effects of early adverse experiences as well as discrimination between the effects of individual exposures. In some cases, longitudinal animal studies can also provide data to test the sensitive-period hypothesis, which posits that early-life social adversity affects later-life health in a manner that is only partially modifiable by later-life experience (105). Strong tests of this hypothesis are difficult to conduct in humans because exposure to early adversity tends to be correlated with later-life exposure to adversity (for example, because of limited social mobility) (106). In animal societies, however, social conditions in adulthood are not always well predicted by social conditions in early life or intergenerationally (99). This decoupling has been leveraged in baboons to show that early adversity in one generation predicts reduced juvenile survival in the next, independently of the juvenile's own early-life experience (107).

Last, studies in animals support the hypothesis that the effects of early adversity on life span among humans are not fully explained by

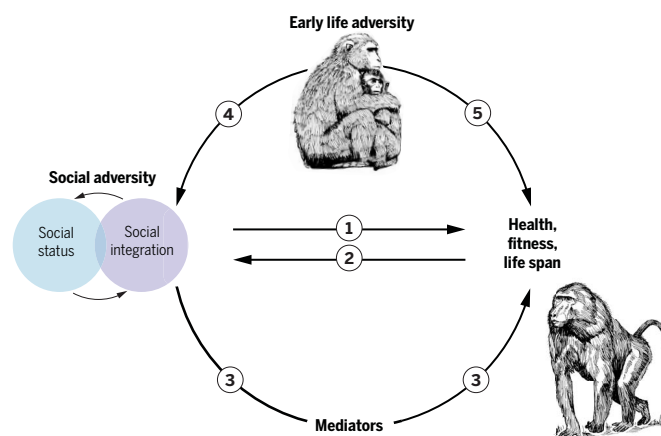


Fig. 4. Pathways linking social factors to health in nonhuman primates.

Box 1. Multiple pathways link social factors to health: Evidence from nonhuman primates.

In humans, the social environment is influenced by a complex set of factors, including income, education, occupation, social prestige, and larger cultural and institutional structures. As defined by the World Health Organization, the social determinants of health are “shaped by the distribution of money, power and resources at global, national and local levels” (188). Social status and social integration also intersect with, and can be influenced by, other social identities, such as race, ethnicity, and gender. By comparison, social environments in nonhuman animals are much simpler and are best studied—and probably most relevant to health, reproduction, and survival—at the local level, where coresident individuals directly interact. Social hierarchies can thus often be summarized by using single-dimensional measures (189).

Nevertheless, as in humans, multiple pathways connect social factors to health and Darwinian fitness in other animals. Several of these pathways are analogous to those developed for human populations (16, 190–192). Social causation (Fig. 4, arrow 1) is strongly supported by studies that manipulate exposure to chronic social stress while holding other aspects of the environment constant (131, 132). By contrast, in species for which social status is determined by physical competition, changes in body condition and physiological measures of endocrine and immune function can precede changes in status (“health selection”) (Fig. 4, arrow 2) (84, 193, 194). Social environmental links to life span can also be mediated through other environmental exposures (Fig. 4, arrow 3). For example, by influencing huddling behavior, social integration affects winter thermoregulation in Barbary macaques (195). Last, early-life adversity can generate social gradients in adulthood (Fig. 4, arrows 4 and 5). In wild female baboons, for example, early maternal loss predicts reduced social integration in adulthood, lower-than-expected adult social status, and shortened life span (85, 99).

As in humans, social status and social relationships can also be interrelated in complex ways (Fig. 4, blue and purple circles). Social status can be relatively independent from social integration, as is the case among wild female baboons (48). Alternatively, social status can structure affiliative social relationships (48, 196, 197); in these cases, high status usually predicts increased social integration, and evidence from captive primates indicates that the effects of status on health-related outcomes may be mediated in part by a path through increased integration (131). Last, developing supportive social relationships can predict subsequent changes in social status. For example, male Assamese macaques that formed stronger social bonds with other males subsequently rose in the dominance hierarchy and also fathered more young (198).

health care access or health risk behaviors such as smoking, alcoholism, or illicit drug use (because these are distinctly human variables). Instead, these studies highlight alternative mechanisms with potential relevance to human studies. For example, female baboons who experienced high levels of early adversity also tend to be more socially isolated from other females later in life (99). In parallel, orphaned elephants have reduced social contact with high-quality social partners (mature adults) compared with nonorphans (108). Given the strong association between affiliative social relationships and mortality risk, these observations suggest that early social adversity may influence later-life outcomes in part through patterning social interactions in adulthood. Such a model is reminiscent of the pathway model proposed for humans: that childhood circumstances affect adult health risk indirectly by putting individuals on trajectories that structure future exposure to later adversity (87, 109).

Biological pathways from social adversity to health

Cross-species comparisons thus suggest that social environments, both in early life and adulthood, are key determinants of life span variation in humans and other social mammals. These parallel findings point to opportunities to draw on data from other social mammals to address outstanding questions about the social determinants of health in humans. Animal models for social gradients in human health

can (i) reduce the complexity of human social environments; (ii) open the door to prospective and intergenerational study designs that can be executed on a much faster time scale; and (iii) in some cases, allow for direct experimental manipulation and invasive sampling. Below, we focus on several emerging themes that draw on one or more of these features, beginning with links between social adversity and health across the life course, and then potential proximate (mechanistic) and ultimate (evolutionary) pathways that could account for these observations. Because the relevant literature is broad, we also point interested readers to taxon- and discipline-specific perspectives reviewed elsewhere (33, 110–114).

Social adversity and health outcomes across the life course

Studies that relate the social environment to health in humans are larger, better replicated, and more representative than their counterparts in any other species (although studies that investigate non-Western nations are still lacking and may influence the types of social conditions classified as “adverse”). However, because they are largely correlational, questions about causal direction persist that can only be partially addressed by using longitudinal or cohort designs (115). One of the most important contributions of studies of social adversity in other social mammals therefore stems from their interpretive clarity, especially in cases in which the social environment itself can be manipulated in controlled experiments: the “gold standard” for inferring causality.

Such studies have long supported a role for social causation, not only for physiological changes that are precursors to disease but also for disease outcomes themselves. For example, low social status more than doubles the rate of coronary artery atherosclerosis and hyperinsulinemia in diet-controlled female long-tailed (cynomolgus) macaques (116, 117). In males, low status and/or social instability also predicts increased prevalence of coronary artery stenosis, and low status (but not social instability) increases susceptibility to experimentally administered adenovirus (118, 119). Relevant to cancer outcomes, lower levels of social reciprocity in female rats predict both earlier tumor onset and shortened life span, and social isolation leads to a 30-fold increase in primary tumor metastasis in mice (120). Thus, manipulation of the social environment in captivity recapitulates social gradients in the leading causes of death in humans, including heart disease, diabetes, and respiratory infections (121).

However, these studies have been short term, relying on genetically predisposed strains or environmental manipulations to accelerate disease outcomes. Only recently have animal studies attempted to model the pattern ob-

served in humans: social gradients that lead to poorer health or elevated mortality from multiple causes, manifested over the life course. In one case, researchers aggregated almost a decade’s worth of data to demonstrate that rhesus macaques randomized into an early maternal loss treatment experienced poorer health later in life, despite standardized housing conditions in adulthood (122). Similarly broad effects have been observed in an experimental mouse model of social status. In the first study to investigate the consequences of chronic stress across natural life spans, persistent exposure to socially dominant animals was shown to shorten the median life spans of socially subordinate male mice by 12.4% (12), an effect size comparable with that of dietary restriction in the same strains (123). Low-status animals also experienced earlier onset of multiorgan lesions, including tumors. In a subset of 17-month-old mice, subordinates had elevated p53 and p16^{Ink4a} markers of cellular senescence and, remarkably, 50% prevalence of early-stage atherosclerotic lesions, which generally occur only in genetically predisposed strains exposed to highly atherogenic diets. By contrast, no lesions were observed in dominant mice.

Replication of these findings will be crucial for assessing their generalizability. Nevertheless, they strongly support the idea that chronic social stress can be sufficiently toxic to explain multiple pathological outcomes, including accelerated senescence (124). In the mouse life span study, for instance, subordinates were housed in proximity to, but physically separated from, dominant mice (12). However, subordinates exhibited close to a twofold increase in glucocorticoid levels, suggesting that simple exposure to threat from an aggressive social partner can induce broad physiological changes (12).

Molecular signatures of social adversity

If social causation contributes to the relationship between social adversity, health, and mortality risk, what are the physiological and molecular changes that mediate this relationship? Efforts to identify these mechanisms have historically focused on neuroendocrine signaling, particularly the contribution of the hypothalamic-pituitary-adrenal (HPA) axis and the sympathetic nervous system (73, 125). Experimental animal models generally support the idea that these pathways are altered by social adversity-induced stress (73, 125, 126), with some corroborating evidence from studies in wild mammals (83, 127). However, the purpose of social adversity-associated changes in neuroendocrine signaling is to communicate a threat to, and regulate the return to, physiological homeostasis. To explain pathophysiology (for example, as a consequence of chronic signaling) (126), social adversity-associated

changes must also lead to changes in target cells and tissues. Understanding how social adversity connects to molecular changes within the cell has become an increasing focus of research, building on a broader sociogenomics literature that shows that social interactions can substantially alter gene regulation (128–130).

Thus far, we know the most about social adversity and gene regulation in peripheral blood cells, which are the most commonly collected sample type in humans and other social mammals. These studies yield a rapidly developing picture of how social adversity causally alters the regulation of the immune system in experimental animal models (131, 132). The most consistent finding from experimental manipulations of the social environment in nonhuman animals is that increased social adversity drives increased expression of genes linked to inflammation, including those that regulate, code for, or interact with biomarkers of chronic stress [such as interleukin-6 (IL6) and IL-1 β]. These changes appear to be shaped by socially patterned differences in the use of immune defense-modulating transcription factors, especially nuclear factor κ B (NF κ B), a master regulator of inflammation (131). In animal models of early social adversity and social status, predicted DNA binding sites for NF κ B are enriched near genes that are more transcriptionally active in socially stressed individuals (131). Further, in rhesus macaques, regions of the genome that are more physically accessible to transcription factor binding in low-status animals also tend to contain NF κ B-binding sites (133). Because NF κ B can be prevented from interacting with DNA through glucocorticoid signaling, this observation suggests a link between functional genomic studies and previous work on stress neuroendocrinology (125). Glucocorticoid resistance—a hallmark of chronic stress—is also associated with increased expression of proinflammatory transcription factors (126).

These patterns parallel those observed in research on social adversity in humans. Although studies in human populations are necessarily correlational, the animal-model work suggests that socially induced stressors are also likely to causally alter gene regulation and the HPA axis in our own species. A growing body of research supports a link between exposure to social adversity and DNA methylation and gene expression markers associated with glucocorticoid signaling and inflammation (134–136). If so, gene regulatory signatures of social adversity may be broadly conserved in social mammals (137). Because relatively few species have been studied at this point, this hypothesis requires data from a broader range of species to test; even in the species studied thus far, often only one sex has been well characterized. Nevertheless, it is notable that cross-species analyses of other aspects of social behavior,

such as territorial aggression and social monogamy, have identified conserved roles for the same gene regulatory pathways in taxa as diverse as rodents, songbirds, frogs, and fish (128, 129).

At the same time, social environmental effects on gene regulation are also context dependent. For example, in rhesus macaque females, the effects of experimentally manipulated social status are magnified in immune cells after exposure to lipopolysaccharide, which stimulates the innate immune response against bacteria (137). Consistent with correlative studies in humans, low-status animals up-regulate pro-inflammatory, NF κ B-regulated pathways relative to high-status animals, whereas high status predicts higher expression of antiviral genes. This pattern has been interpreted as a stress-mediated trade-off between antibacterial and antiviral defense (137). However, recent work indicates that this pattern is contingent on the local cellular environment: Key regulators of antiviral defense that are positively correlated with social status after exposure to bacterial compounds actually become negatively correlated with social status in the same animals, after challenge with a viral mimic (138). Such studies may provide a window into understanding why the effects of social adversity differ across settings and into the basis of cumulative risk and multiple hit models (102, 103). However, they also caution against the idea that there is a simple map between social environmental effects on immune gene expression and differential susceptibility to specific pathogens.

Evolutionary frameworks for the social determinants of health

The studies above focus on the proximate physiological and molecular mechanisms that explain the social determinants of health. However, the congruence between findings in humans and observations in other social mammals not only suggests that nonhuman species can serve as effective models for humans but also that social gradients in health may be coupled with the evolution of social living itself. Comparative studies can therefore also contribute by highlighting the evolutionary logic that explains social gradients (83, 139, 140). Such studies have already been key to understanding the evolutionary costs and benefits of transitions to group living (55, 141, 142).

Social gradients within species arise because social costs and benefits are not equally distributed across individuals coexisting in the same social group. Consistent, species-level differences in the steepness of social hierarchies and the stability of social bonds emerge from the need to resolve this tension, as discussed in a long history of comparative work on the emergence of egalitarian versus “despotic” animal societies (140, 143). However, individu-

als are likely subject to additional selection for sensitivity to the quality of social relationships within social groups (144). For example, the concept of the “dominance behavioral system,” developed in evolutionary psychology, argues that humans and other social animals have evolved finely tuned biological sensors to evaluate their and others’ relative social status (145). In support of this argument, work in mice has identified specific sensory and neural substrates for assessing dominance and social integration (146–149). However, we know of no case to date in which the fitness consequences of variation in social sensitivity has been evaluated in a natural social mammal population. Doing so would require measuring interindividual differences in the response to a common social environment, accurately assessing the “appropriate” social response, and potentially measuring subjective social experience. The increasing availability of life course data from wild mammals as well as new methods for quantifying perceived social stress (such as in captive rhesus macaques) (150) may make such studies feasible in the near future.

By contrast, data from wild social mammals have already brought clarity to evolutionary hypotheses about the long-term health effects of early adversity. For example, an extensive body of theory has been developed to account for observations of such effects in humans (151–154). The most commonly invoked ideas focus on predictive adaptive responses (PARs), which propose that early-life effects evolved because natural selection favors organisms that tailor their later-life phenotype to the environmental cues they experience in early life. PAR models argue that it is the mismatch between early adverse conditions and later, more benign conditions that produces the adverse health effects of early adversity. However, because predictive models assume that early-life environmental cues must be reliable indicators of the later-life environment, theoretical work suggests that PARs are unlikely to evolve in long-lived species (155). In nonhuman animals, the best empirical support for PARs comes from short-lived species (156, 157). By contrast, studies in long-lived mammals provide better support for an alternative set of models: developmental constraints (158–161). Developmental constraints models posit that early-life effects evolve because they allow immediate survival, at the expense of optimal development, even if they incur later-life costs; they are the result of natural selection on the ability to “make the best of a bad situation.” If so, individuals who experienced early adversity may perform quite poorly when faced with adverse environments in adulthood—a conclusion with substantially different intervention and policy implications than those of the PAR model.

Tests of more refined PAR models are ongoing (162, 163). However, the above work

already illustrates the value of studies in nonhuman species for testing evolutionary arguments relevant to social gradients in health (164, 165). It also highlights the challenges in clearly discriminating adaptive from non-adaptive responses: Apparently costly responses to social adversity can be favored by natural selection if they are better than no phenotypic adjustment at all (166).

Conclusions and new directions

The available evidence indicates that social impacts on life span are a shared phenomenon across humans and other social mammals and that the health-related outcomes of social adversity in nonhuman animals parallel socially patterned pathologies in humans. To some degree, the mechanisms that underlie these observations are also similar across species: Social conditions that promote chronic stress also predict increased inflammation, HPA axis dysregulation, and changes to sympathetic nervous system signaling (126). These findings suggest a shared biology underlying the influence of social gradients and a coherent evolutionary logic for when these gradients tend to be shallower versus steeper—arguments that have been made in various forms over the years (32). Only recently, however, have they been supported by both experimental tests for causal outcomes and data on natural mortality, with correspondingly refined estimates from very large studies in humans (4, 12, 99, 131).

A shared biology in turn suggests that integrating human and nonhuman animal studies can help address longstanding questions about the social determinants of health. Research at this interface should open several new opportunities. First, the findings outlined here argue that the social determinants of health should be of central interest to biologists as well as social scientists. This is not yet the case for many disciplines; for example, the field of genomics was recently taken to task for ignoring the literature on social gradients in health and, as a consequence, redefining health disparities in terms of population genetic diversity (a genetic explanation) instead of recognizing its fundamental origins in the social environment (167). Research with natural links to the social determinants of health has been similarly limited in other disciplines; for example, recent studies that compare genetic and nonheritable predictors of immune function consider age, sex, and past pathogen exposure as environmental factors but not the social environment (168). Broadening this perspective presents an opportunity to leverage new methodological advances to understand the causes and consequences of social gradients, including scope for potential intervention. Animal model studies may be ideal for testing proposed interventions because they ensure compliance and eliminate other confounding factors.

Second, the parallels between studies highlight untapped opportunities to translate biological outcome measures across fields, especially molecular and physiological markers of social adversity and health. One important gap to fill involves the fact that nearly all of the evidence that social adversity compromises natural life span in social mammals comes from natural populations. By contrast, the best evidence for social causation of specific physiological or health outcomes comes from laboratory studies. Demonstrating that such findings are not artifacts of captivity—for example, by translating these outcome measures to natural populations—is crucial for understanding whether the relationship between social adversity and life span in nature can be explained, at least in part, by the mechanisms being identified in experimental studies. For example, although the prevailing model for social causation in laboratory studies invokes exposure to chronic social stress, some researchers have argued that animals in their natural environments are unlikely to experience chronic stress, or at least not to the degree that it could shorten life span (169).

Last, researchers must expand the set of study systems to other species and tissue types (especially the brain) and to a more diverse set of human populations. Increased diversity will help reveal how variation in social gradients emerge. For example, differences in the routes through which status is attained, the steepness and regularity of hierarchy enforcement, and the availability of coping outlets have all been proposed to modify the severity of social gradients (1, 32, 139). In humans and at least six other primates, increased life span equality is positively correlated with increased life expectancy overall, in support of the idea that members of more egalitarian groups tend to have longer survival (170, 171). In some species, the canonical direction of social gradients may also be reversed. In species in which competition for high status is energetically demanding, as it is in hierarchies that are based on physical competition (83, 127), high-ranking individuals have been shown to exhibit higher glucocorticoid levels, up-regulate inflammation-related pathways, and experience accelerated “biological aging” (based on telomere shortening and epigenetic clock prediction) (79, 84, 127, 172). Such results stress that different types of social systems can produce different kinds of gradients. Understanding why—for example, by use of evolutionary comparative methods across species—may suggest ways to decouple social environmental variation from its negative health consequences in humans.

REFERENCES AND NOTES

- R. M. Sapolsky, Social status and health in humans and other animals. *Annu. Rev. Anthropol.* **33**, 393–418 (2004). doi: [10.1146/annurev.anthro.33.070203.144000](https://doi.org/10.1146/annurev.anthro.33.070203.144000)
- R. Chetty et al., The association between income and life expectancy in the United States, 2001–2014. *JAMA* **315**, 1750–1766 (2016). doi: [10.1001/jama.2016.4226](https://doi.org/10.1001/jama.2016.4226); pmid: [27063997](https://pubmed.ncbi.nlm.nih.gov/27063997/)
- S. Stringhini et al., Socioeconomic status and the 25 × 25 risk factors as determinants of premature mortality: A multicohort study and meta-analysis of 1.7 million men and women. *Lancet* **389**, 1229–1237 (2017). doi: [10.1016/S0140-6736\(16\)32380-7](https://doi.org/10.1016/S0140-6736(16)32380-7); pmid: [28159391](https://pubmed.ncbi.nlm.nih.gov/28159391/)
- J. Holt-Lunstad, T. B. Smith, J. B. Layton, Social relationships and mortality risk: A meta-analytic review. *PLOS Med.* **7**, e1000316 (2010). doi: [10.1371/journal.pmed.1000316](https://doi.org/10.1371/journal.pmed.1000316); pmid: [20668659](https://pubmed.ncbi.nlm.nih.gov/20668659/)
- M. Marmot, Health equity in England: The Marmot review 10 years on. *BMJ* **368**, m693 (2020). doi: [10.1136/bmj.m693](https://doi.org/10.1136/bmj.m693); pmid: [32094110](https://pubmed.ncbi.nlm.nih.gov/32094110/)
- A. M. O’Rand, S. M. Lynch, Socioeconomic status, health, and mortality in aging populations, in *Future Directions for the Demography of Aging*, Hayward, K. Majumdar, Eds. (National Academies Press, 2018), pp. 67–95.
- L. C. Hawkey, J. T. Cacioppo, Loneliness matters: A theoretical and empirical review of consequences and mechanisms. *Ann. Behav. Med.* **40**, 218–227 (2010). doi: [10.1007/s12160-010-9210-8](https://doi.org/10.1007/s12160-010-9210-8); pmid: [20652462](https://pubmed.ncbi.nlm.nih.gov/20652462/)
- J. Holt-Lunstad, T. B. Smith, M. Baker, T. Harris, D. Stephenson, Loneliness and social isolation as risk factors for mortality: A meta-analytic review. *Perspect. Psychol. Sci.* **10**, 227–237 (2015). doi: [10.1177/1745691614568352](https://doi.org/10.1177/1745691614568352); pmid: [25910392](https://pubmed.ncbi.nlm.nih.gov/25910392/)
- V. J. Felitti et al., Relationship of childhood abuse and household dysfunction to many of the leading causes of death in adults. The Adverse Childhood Experiences (ACE) Study. *Am. J. Prev. Med.* **14**, 245–258 (1998). doi: [10.1016/S0749-3797\(98\)00017-8](https://doi.org/10.1016/S0749-3797(98)00017-8); pmid: [9635069](https://pubmed.ncbi.nlm.nih.gov/9635069/)
- Y. C. Yang, K. Schorpp, C. Boen, M. Johnson, K. M. Harris, Socioeconomic status and biological risks for health and illness across the life course. *J. Gerontol. B Psychol. Sci. Soc. Sci.* **75**, 613–624 (2020). doi: [10.1093/geronb/gby108](https://doi.org/10.1093/geronb/gby108); pmid: [30252104](https://pubmed.ncbi.nlm.nih.gov/30252104/)
- C. A. Shively, M. E. Wilson, *Social Inequalities in Health in Nonhuman Primates: The Biology of the Gradient* (Springer, 2016).
- M. Razzoli et al., Social stress shortens lifespan in mice. *Aging Cell* **17**, e12778 (2018). doi: [10.1111/ace1.12778](https://doi.org/10.1111/ace1.12778); pmid: [29806171](https://pubmed.ncbi.nlm.nih.gov/29806171/)
- P. Muennig, Health selection vs. causation in the income gradient: What can we learn from graphical trends? *J. Health Care Poor Underserved* **19**, 574–579 (2008). doi: [10.1353/hpu.0.0018](https://doi.org/10.1353/hpu.0.0018); pmid: [18469427](https://pubmed.ncbi.nlm.nih.gov/18469427/)
- H. Kröger, E. Pakpahan, R. Hoffmann, What causes health inequality? A systematic review on the relative importance of social causation and health selection. *Eur. J. Public Health* **25**, 951–960 (2015). doi: [10.1093/eurpub/ckv111](https://doi.org/10.1093/eurpub/ckv111); pmid: [26089181](https://pubmed.ncbi.nlm.nih.gov/26089181/)
- R. Hoffmann, H. Kröger, S. Geyer, Social causation versus health selection in the life course: Does their relative importance differ by dimension of SES? *Soc. Indic. Res.* **141**, 1341–1367 (2019). doi: [10.1007/s12050-018-1871-x](https://doi.org/10.1007/s12050-018-1871-x)
- J. B. Kane, K. M. Harris, S. P. Morgan, D. K. Guilkey, Pathways of health and human capital from adolescence into young adulthood. *Soc. Forces* **96**, 949–976 (2018). doi: [10.1093/sf/sox079](https://doi.org/10.1093/sf/sox079); pmid: [30555185](https://pubmed.ncbi.nlm.nih.gov/30555185/)
- M. Anderson, M. Marmot, The effects of promotions on heart disease: Evidence from Whitehall. *Econ. J. (Lond.)* **122**, 555–589 (2012). doi: [10.1111/j.1468-0297.2011.02472.x](https://doi.org/10.1111/j.1468-0297.2011.02472.x)
- S. Stringhini et al., Health behaviours, socioeconomic status, and mortality: Further analyses of the British Whitehall II and the French GAZEL prospective cohorts. *PLOS Med.* **8**, e1000419 (2011). doi: [10.1371/journal.pmed.1000419](https://doi.org/10.1371/journal.pmed.1000419); pmid: [21364974](https://pubmed.ncbi.nlm.nih.gov/21364974/)
- A. Case, C. Paxson, The long reach of childhood health and circumstance: Evidence from the Whitehall II study. *Econ. J.* **121**, F183–F204 (2011). doi: [10.1111/j.1468-0297.2011.02447.x](https://doi.org/10.1111/j.1468-0297.2011.02447.x); pmid: [22228912](https://pubmed.ncbi.nlm.nih.gov/22228912/)
- J. B. Kane, K. M. Harris, A. M. Siega-Riz, Intergenerational pathways linking maternal early life adversity to offspring birthweight. *Soc. Sci. Med.* **207**, 89–96 (2018). doi: [10.1016/j.socscimed.2018.04.049](https://doi.org/10.1016/j.socscimed.2018.04.049); pmid: [29734059](https://pubmed.ncbi.nlm.nih.gov/29734059/)
- S. Sealy-Jefferson, C. Giurgescu, L. Helmkamp, D. P. Misra, T. L. Osypuk, Perceived physical and social residential environment and preterm delivery in African-American women. *Am. J. Epidemiol.* **182**, 485–493 (2015). doi: [10.1093/aje/kwv106](https://doi.org/10.1093/aje/kwv106); pmid: [26163532](https://pubmed.ncbi.nlm.nih.gov/26163532/)
- S. Harper, E. C. Strumpf, Social epidemiology: Questionable answers and answerable questions. *Epidemiology* **23**, 795–798 (2012). doi: [10.1097/EDE.0b013e31826d078d](https://doi.org/10.1097/EDE.0b013e31826d078d); pmid: [23038109](https://pubmed.ncbi.nlm.nih.gov/23038109/)
- E. J. Ozer, L. C. H. Fernald, A. Weber, E. P. Flynn, T. J. VanderWeele, Does alleviating poverty affect mothers’ depressive symptoms? A quasi-experimental investigation of Mexico’s Oportunidades programme. *Int. J. Epidemiol.* **40**, 1565–1576 (2011). doi: [10.1093/ije/dyr103](https://doi.org/10.1093/ije/dyr103); pmid: [21737404](https://pubmed.ncbi.nlm.nih.gov/21737404/)
- E. J. Costello, A. Erkanli, W. Copeland, A. Angold, Association of family income supplements in adolescence with development of psychiatric and substance use disorders in adulthood among an American Indian population. *JAMA* **303**, 1954–1960 (2010). doi: [10.1001/jama.2010.621](https://doi.org/10.1001/jama.2010.621); pmid: [20483972](https://pubmed.ncbi.nlm.nih.gov/20483972/)
- R. Hamad, D. H. Rehkopf, Poverty and child development: A longitudinal study of the impact of the earned income tax credit. *Am. J. Epidemiol.* **183**, 775–784 (2016). doi: [10.1093/aje/kwv317](https://doi.org/10.1093/aje/kwv317); pmid: [27056961](https://pubmed.ncbi.nlm.nih.gov/27056961/)
- J. Cassel, The contribution of the social environment to host resistance: The Fourth Wade Hampton Frost Lecture. *Am. J. Epidemiol.* **104**, 107–123 (1976). doi: [10.1093/oxfordjournals.aje.a112281](https://doi.org/10.1093/oxfordjournals.aje.a112281); pmid: [782233](https://pubmed.ncbi.nlm.nih.gov/782233/)
- R. Koch, An address on bacteriological research. *BMJ* **2**, 380–383 (1890). doi: [10.1136/bmj.2.1546.380](https://doi.org/10.1136/bmj.2.1546.380); pmid: [20753110](https://pubmed.ncbi.nlm.nih.gov/20753110/)
- B. G. Link, J. Phelan, Social conditions as fundamental causes of disease. *J. Health Soc. Behav.* **35**, 80–94 (1995). doi: [10.2307/2626958](https://doi.org/10.2307/2626958); pmid: [7560851](https://pubmed.ncbi.nlm.nih.gov/7560851/)
- B. S. McEwen, T. Seeman, Protective and damaging effects of mediators of stress. Elaborating and testing the concepts of allostasis and allostatic load. *Ann. N. Y. Acad. Sci.* **896**, 30–47 (1999). doi: [10.1111/j.1749-6632.1999.tb08103.x](https://doi.org/10.1111/j.1749-6632.1999.tb08103.x); pmid: [10681886](https://pubmed.ncbi.nlm.nih.gov/10681886/)
- C. Hertzman, T. Boyce, How experience gets under the skin to create gradients in developmental health. *Annu. Rev. Public Health* **31**, 329–347, 3p, 347 (2010). doi: [10.1146/annurev.publhealth.012809.103538](https://doi.org/10.1146/annurev.publhealth.012809.103538); pmid: [20070189](https://pubmed.ncbi.nlm.nih.gov/20070189/)
- Y. C. Yang et al., Social relationships and physiological determinants of longevity across the human life span. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 578–583 (2016). doi: [10.1073/pnas.1511085112](https://doi.org/10.1073/pnas.1511085112); pmid: [26729882](https://pubmed.ncbi.nlm.nih.gov/26729882/)
- R. M. Sapolsky, The influence of social hierarchy on primate health. *Science* **308**, 648–652 (2005). doi: [10.1126/science.1106477](https://doi.org/10.1126/science.1106477); pmid: [15860617](https://pubmed.ncbi.nlm.nih.gov/15860617/)
- S. A. Cavigelli, M. J. Caruso, Sex, social status and physiological stress in primates: The importance of social and glucocorticoid dynamics. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **370**, 20140103 (2015). doi: [10.1098/rstb.2014.0103](https://doi.org/10.1098/rstb.2014.0103); pmid: [25870390](https://pubmed.ncbi.nlm.nih.gov/25870390/)
- I. T. Elo, Social class differentials in health and mortality: Patterns and explanations in comparative perspective. *Annu. Rev. Sociol.* **35**, 553–572 (2009). doi: [10.1146/annurev-soc-070308-115929](https://doi.org/10.1146/annurev-soc-070308-115929)
- K. M. Harris, K. M. Schorpp, Integrating biomarkers in social stratification and health research. *Annu. Rev. Sociol.* **44**, 361–386 (2018). doi: [10.1146/annurev-soc-060116-053339](https://doi.org/10.1146/annurev-soc-060116-053339); pmid: [30918418](https://pubmed.ncbi.nlm.nih.gov/30918418/)
- J. Holt-Lunstad, Why social relationships are important for physical health: A systems approach to understanding and modifying risk and protection. *Annu. Rev. Psychol.* **69**, 437–458 (2018). doi: [10.1146/annurev-psych-122116-011902](https://doi.org/10.1146/annurev-psych-122116-011902); pmid: [29035688](https://pubmed.ncbi.nlm.nih.gov/29035688/)
- B. Johnson, J. R. Carey, Hierarchy and connectedness as determinants of health and longevity in social insects, in *Sociality, Hierarchy, Health: Comparative Biodemography: A Collection of Papers* (National Academies Press, 2014), pp. 269–293.
- B. S. Rowntree, *Poverty: A Study of Town Life* (Macmillan, 1901).
- C. Darwin, *On The Origin of Species by Means of Natural Selection, or Preservation of Favoured Races in the Struggle for Life* (John Murray, 1859).
- R. I. M. Dunbar, S. Shultz, Evolution in the social brain. *Science* **317**, 1344–1347 (2007). doi: [10.1126/science.1145463](https://doi.org/10.1126/science.1145463); pmid: [17823343](https://pubmed.ncbi.nlm.nih.gov/17823343/)
- J. B. Silk, The adaptive value of sociality in mammalian groups. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **362**, 539–559 (2007). doi: [10.1098/rstb.2006.1994](https://doi.org/10.1098/rstb.2006.1994); pmid: [17363359](https://pubmed.ncbi.nlm.nih.gov/17363359/)
- T. H. Clutton-Brock, *Reproductive Success: Studies of Individual Variation in Contrasting Breeding Systems* (Univ. Chicago Press, 1988).
- L. F. Berkman, T. Glass, I. Brissette, T. E. Seeman, From social integration to health: Durkheim in the new millennium. *Soc. Sci. Med.* **51**, 843–857 (2000). doi: [10.1016/S0277-9536\(00\)00065-4](https://doi.org/10.1016/S0277-9536(00)00065-4); pmid: [10972429](https://pubmed.ncbi.nlm.nih.gov/10972429/)

44. A. C. Markham, S. C. Alberts, J. Altmann, Intergroup conflict: Ecological predictors of winning and consequences of defeat in a wild primate population. *Anim. Behav.* **82**, 399–403 (2012). doi: [10.1016/j.anbehav.2012.05.009](https://doi.org/10.1016/j.anbehav.2012.05.009); pmid: [22837555](https://pubmed.ncbi.nlm.nih.gov/22837555/)
45. J. S. House, K. R. Landis, D. Umberson, Social relationships and health. *Science* **241**, 540–545 (1988). doi: [10.1126/science.3399889](https://doi.org/10.1126/science.3399889); pmid: [3399889](https://pubmed.ncbi.nlm.nih.gov/3399889/)
46. L. F. Berkman, S. L. Syme, Social networks, host resistance, and mortality: A nine-year follow-up study of Alameda County residents. *Am. J. Epidemiol.* **109**, 186–204 (1979). doi: [10.1093/oxfordjournals.aje.a112674](https://doi.org/10.1093/oxfordjournals.aje.a112674); pmid: [425958](https://pubmed.ncbi.nlm.nih.gov/425958/)
47. J. B. Silk et al., Strong and consistent social bonds enhance the longevity of female baboons. *Curr. Biol.* **20**, 1359–1361 (2010). doi: [10.1016/j.cub.2010.05.067](https://doi.org/10.1016/j.cub.2010.05.067); pmid: [20598541](https://pubmed.ncbi.nlm.nih.gov/20598541/)
48. E. A. Archie, J. Tung, M. Clark, J. Altmann, S. C. Alberts, Social affiliation matters: Both same-sex and opposite-sex relationships predict survival in wild female baboons. *Proc. Biol. Sci.* **281**, 20141261 (2014). doi: [10.1098/rspb.2014.1261](https://doi.org/10.1098/rspb.2014.1261); pmid: [25209936](https://pubmed.ncbi.nlm.nih.gov/25209936/)
49. M. A. Stanton, J. Mann, Early social networks predict survival in wild bottlenose dolphins. *PLOS ONE* **7**, e47508 (2012). doi: [10.1371/journal.pone.0047508](https://doi.org/10.1371/journal.pone.0047508); pmid: [23077627](https://pubmed.ncbi.nlm.nih.gov/23077627/)
50. C. M. V. Nunez, J. S. Adelman, D. I. Rubenstein, Sociality increases juvenile survival after a catastrophic event in the feral horse (*Equus caballus*). *Behav. Ecol.* **26**, 138–147 (2015). doi: [10.1093/beheco/aru163](https://doi.org/10.1093/beheco/aru163)
51. B. Dantzer et al., Density triggers maternal hormones that increase adaptive offspring growth in a wild mammal. *Science* **340**, 1215–1217 (2013). doi: [10.1126/science.1235765](https://doi.org/10.1126/science.1235765); pmid: [23599265](https://pubmed.ncbi.nlm.nih.gov/23599265/)
52. D. T. Blumstein, D. M. Williams, A. N. Lim, S. Kroeger, J. G. A. Martin, Strong social relationships are associated with decreased longevity in a facultatively social mammal. *Proc. Biol. Sci.* **285**, 20171934 (2018). doi: [10.1098/rspb.2017.1934](https://doi.org/10.1098/rspb.2017.1934); pmid: [29343594](https://pubmed.ncbi.nlm.nih.gov/29343594/)
53. A. P. Montero, D. M. Williams, J. G. A. Martin, D. T. Blumstein, More social female yellow-bellied marmots, *Marmota flaviventris*, have enhanced summer survival. *Anim. Behav.* **160**, 113–119 (2020). doi: [10.1016/j.anbehav.2019.12.013](https://doi.org/10.1016/j.anbehav.2019.12.013)
54. N. A. Thompson, M. Cords, Stronger social bonds do not always predict greater longevity in a gregarious primate. *Ecol. Evol.* **8**, 1604–1614 (2018). doi: [10.1002/ece3.3781](https://doi.org/10.1002/ece3.3781); pmid: [29435236](https://pubmed.ncbi.nlm.nih.gov/29435236/)
55. D. Lukas, T. H. Clutton-Brock, The evolution of social monogamy in mammals. *Science* **341**, 526–530 (2013). doi: [10.1126/science.1238677](https://doi.org/10.1126/science.1238677); pmid: [23896459](https://pubmed.ncbi.nlm.nih.gov/23896459/)
56. Y. C. Yang, M. K. McClintock, M. Kozloski, T. Li, Social isolation and adult mortality: The role of chronic inflammation and sex differences. *J. Health Soc. Behav.* **54**, 183–203 (2013). doi: [10.1177/0022146513485244](https://doi.org/10.1177/0022146513485244); pmid: [23653312](https://pubmed.ncbi.nlm.nih.gov/23653312/)
57. C. E. Boen et al., Social relationships, inflammation, and cancer survival. *Cancer Epidemiol. Biomarkers Prev.* **27**, 541–549 (2018). doi: [10.1158/1055-9965.EPI-17-0836](https://doi.org/10.1158/1055-9965.EPI-17-0836); pmid: [29475966](https://pubmed.ncbi.nlm.nih.gov/29475966/)
58. L. A. D. Campbell, P. J. Tkaczynski, J. Lehmann, M. Mouna, B. Majolo, Social thermoregulation as a potential mechanism linking sociality and fitness: Barbary macaques with more social partners form larger huddles. *Sci. Rep.* **8**, 6074 (2018). doi: [10.1038/s41598-018-24373-4](https://doi.org/10.1038/s41598-018-24373-4); pmid: [29666428](https://pubmed.ncbi.nlm.nih.gov/29666428/)
59. R. McFarland et al., Social integration confers thermal benefits in a gregarious primate. *J. Anim. Ecol.* **84**, 871–878 (2015). doi: [10.1111/1365-2656.12329](https://doi.org/10.1111/1365-2656.12329); pmid: [25581128](https://pubmed.ncbi.nlm.nih.gov/25581128/)
60. E. Vander Wal, M. Festa-Bianchet, D. Réale, D. W. Coltman, F. Pelletier, Sex-based differences in the adaptive value of social behavior contrasted against morphology and environment. *Ecology* **96**, 631–641 (2015). doi: [10.1890/14-1320.1](https://doi.org/10.1890/14-1320.1); pmid: [26236860](https://pubmed.ncbi.nlm.nih.gov/26236860/)
61. S. Ellis et al., Mortality risk and social network position in resident killer whales: Sex differences and the importance of resource abundance. *Proc. Biol. Sci.* **284**, 20171313 (2017). doi: [10.1098/rspb.2017.1313](https://doi.org/10.1098/rspb.2017.1313); pmid: [29070720](https://pubmed.ncbi.nlm.nih.gov/29070720/)
62. D. R. Farine, H. Whitehead, Constructing, conducting and interpreting animal social network analysis. *J. Anim. Ecol.* **84**, 1144–1163 (2015). doi: [10.1111/1365-2656.12418](https://doi.org/10.1111/1365-2656.12418); pmid: [26172345](https://pubmed.ncbi.nlm.nih.gov/26172345/)
63. D. L. Cheney, J. B. Silk, R. M. Seyfarth, Network connections, dyadic bonds and fitness in wild female baboons. *R. Soc. Open Sci.* **3**, 160255 (2016). doi: [10.1098/rsos.160255](https://doi.org/10.1098/rsos.160255); pmid: [27493779](https://pubmed.ncbi.nlm.nih.gov/27493779/)
64. J. B. Silk, R. M. Seyfarth, D. L. Cheney, Quality versus quantity: Do weak bonds enhance the fitness of female baboons? *Anim. Behav.* **140**, 207–211 (2018). doi: [10.1016/j.anbehav.2018.04.013](https://doi.org/10.1016/j.anbehav.2018.04.013)
65. S. Ellis, N. Snyder-Mackler, A. Ruiz-Lambides, M. L. Platt, L. J. N. Brent, Deconstructing sociality: The types of social connections that predict longevity in a group-living primate. *Proc. Biol. Sci.* **286**, 20191991 (2019). doi: [10.1098/rspb.2019.1991](https://doi.org/10.1098/rspb.2019.1991); pmid: [31822256](https://pubmed.ncbi.nlm.nih.gov/31822256/)
66. A. Hajat, J. S. Kaufman, K. M. Rose, A. Siddiqi, J. C. Thomas, Long-term effects of wealth on mortality and self-rated health status. *Am. J. Epidemiol.* **173**, 192–200 (2011). doi: [10.1093/aje/kwq348](https://doi.org/10.1093/aje/kwq348); pmid: [21059808](https://pubmed.ncbi.nlm.nih.gov/21059808/)
67. R. G. Wilkinson, M. G. Marmot, *Social Determinants of Health: The Solid Facts* (World Health Organization, 2003).
68. M. G. Marmot, A. M. Adelstein, N. Robinson, G. A. Rose, Changing social-class distribution of heart disease. *BMJ* **2**, 1109–1112 (1978). doi: [10.1136/bmj.2.6145.1109](https://doi.org/10.1136/bmj.2.6145.1109); pmid: [709255](https://pubmed.ncbi.nlm.nih.gov/709255/)
69. G. Rose, M. G. Marmot, Social class and coronary heart disease. *Br. Heart J.* **45**, 13–19 (1981). doi: [10.1136/hrt.45.1.13](https://doi.org/10.1136/hrt.45.1.13); pmid: [7459161](https://pubmed.ncbi.nlm.nih.gov/7459161/)
70. M. G. Marmot, R. G. Wilkinson, *Social Determinants of Health* (Oxford Univ. Press, 1999).
71. P. M. Lantz, E. Golberstein, J. S. House, J. Morenoff, Socioeconomic and behavioral risk factors for mortality in a national 19-year prospective study of U.S. adults. *Soc. Sci. Med.* **70**, 1558–1566 (2010). doi: [10.1016/j.socscimed.2010.02.003](https://doi.org/10.1016/j.socscimed.2010.02.003); pmid: [20226579](https://pubmed.ncbi.nlm.nih.gov/20226579/)
72. C. A. Shively, S. M. Day, Social inequalities in health in nonhuman primates. *Neurobiol. Stress* **1**, 156–163 (2014). doi: [10.1016/j.ynstr.2014.11.005](https://doi.org/10.1016/j.ynstr.2014.11.005); pmid: [27589665](https://pubmed.ncbi.nlm.nih.gov/27589665/)
73. J. Kohn, L. Panagiotakopoulos, G. N. Neigh, The effects of social experience on the stress system and immune function in nonhuman primates, in *Developments in Primatology: Progress and Prospects* (2016), pp. 49–77.
74. A. Bartolomucci, Social stress, immune functions and disease in rodents. *Front. Neuroendocrinol.* **28**, 28–49 (2007). doi: [10.1016/j.ynfrne.2007.02.001](https://doi.org/10.1016/j.ynfrne.2007.02.001); pmid: [17379284](https://pubmed.ncbi.nlm.nih.gov/17379284/)
75. W. Goymann, J. C. Wingfield, Allostatic load, social status and stress hormones: The costs of social status matter. *Anim. Behav.* **67**, 591–602 (2004). doi: [10.1016/j.anbehav.2003.08.007](https://doi.org/10.1016/j.anbehav.2003.08.007); pmid: [30858321](https://pubmed.ncbi.nlm.nih.gov/30858321/)
76. D. H. Abbott, A. S. McNeilly, S. F. Lunn, M. J. Hulme, F. J. Burden, Inhibition of ovarian function in subordinate female marmoset monkeys (*Callithrix jacchus jacchus*). *J. Reprod. Fertil.* **63**, 335–345 (1981). doi: [10.1530/jrf.0.0630335](https://doi.org/10.1530/jrf.0.0630335); pmid: [6795346](https://pubmed.ncbi.nlm.nih.gov/6795346/)
77. E. D. Strauss, K. E. Holekamp, Social alliances improve rank and fitness in convention-based societies. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 8919–8924 (2019). doi: [10.1073/pnas.1810384116](https://doi.org/10.1073/pnas.1810384116); pmid: [30858321](https://pubmed.ncbi.nlm.nih.gov/30858321/)
78. D. von Holst, H. Hutzelmeyer, P. Kaetzke, M. Khaschei, R. Schönhöner, Social rank, stress, fitness, and life expectancy in wild rabbits. *Naturwissenschaften* **86**, 388–393 (1999). doi: [10.1007/s001140050638](https://doi.org/10.1007/s001140050638); pmid: [10481826](https://pubmed.ncbi.nlm.nih.gov/10481826/)
79. D. L. Cram et al., Rank-related contrasts in longevity arise from extra-group excursions not delayed senescence in a cooperative mammal. *Curr. Biol.* **28**, 2934–2939.e4 (2018). doi: [10.1016/j.cub.2018.07.021](https://doi.org/10.1016/j.cub.2018.07.021); pmid: [30174185](https://pubmed.ncbi.nlm.nih.gov/30174185/)
80. S. K. Wasser, G. W. Norton, R. J. Rhine, N. Klein, S. Kleindorfer, Ageing and social rank effects on the reproductive system of free-ranging yellow baboons (*Papio cynocephalus*) at Mikumi National Park, Tanzania. *Hum. Reprod. Update* **4**, 430–438 (1998). doi: [10.1093/humupd/4.4.430](https://doi.org/10.1093/humupd/4.4.430); pmid: [9825857](https://pubmed.ncbi.nlm.nih.gov/9825857/)
81. G. E. Blomquist, D. S. Sade, J. D. Berard, Rank-related fitness differences and their demographic pathways in semi-free-ranging rhesus macaques (*Macaca mulatta*). *Int. J. Primatol.* **32**, 193–208 (2011). doi: [10.1007/s10764-010-9461-z](https://doi.org/10.1007/s10764-010-9461-z)
82. M. A. van Noordwijk, C. P. van Schaik, The effects of dominance rank and group size on female lifetime reproductive success in wild long-tailed macaques, *Macaca fascicularis*. *Primates* **40**, 105–130 (1999). doi: [10.1007/BF02557705](https://doi.org/10.1007/BF02557705); pmid: [23179535](https://pubmed.ncbi.nlm.nih.gov/23179535/)
83. D. H. Abbott et al., Are subordinates always stressed? A comparative analysis of rank differences in cortisol levels among primates. *Horm. Behav.* **43**, 67–82 (2003). doi: [10.1016/S0018-506X\(02\)00037-5](https://doi.org/10.1016/S0018-506X(02)00037-5); pmid: [12614636](https://pubmed.ncbi.nlm.nih.gov/12614636/)
84. A. J. Lea et al., Dominance rank-associated gene expression is widespread, sex-specific, and a precursor to high social status in wild male baboons. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E12163–E12171 (2018). doi: [10.1073/pnas.1811967115](https://doi.org/10.1073/pnas.1811967115); pmid: [30538194](https://pubmed.ncbi.nlm.nih.gov/30538194/)
85. A. J. Lea, N. H. Learn, M. J. Theus, J. Altmann, S. C. Alberts, Complex sources of variance in female dominance rank in a nepotistic society. *Anim. Behav.* **94**, 87–99 (2014). doi: [10.1016/j.anbehav.2014.05.019](https://doi.org/10.1016/j.anbehav.2014.05.019); pmid: [26997663](https://pubmed.ncbi.nlm.nih.gov/26997663/)
86. K. E. Holekamp, J. E. Smith, C. C. Strelloff, R. C. Van Horn, H. E. Watts, Society, demography and genetic structure in the spotted hyena. *Mol. Ecol.* **21**, 613–632 (2012). doi: [10.1111/j.1365-294X.2011.05240.x](https://doi.org/10.1111/j.1365-294X.2011.05240.x); pmid: [21880088](https://pubmed.ncbi.nlm.nih.gov/21880088/)
87. M. D. Hayward, B. K. Gorman, The long arm of childhood: The influence of early-life social conditions on men's mortality. *Demography* **41**, 87–107 (2004). doi: [10.1353/dem.2004.0005](https://doi.org/10.1353/dem.2004.0005); pmid: [15074126](https://pubmed.ncbi.nlm.nih.gov/15074126/)
88. K. F. Ferraro, M. H. Schafer, L. R. Wilkinson, Childhood disadvantage and health problems in middle and later life: Early imprints on physical health? *Am. Sociol. Rev.* **81**, 107–133 (2016). doi: [10.1177/0003122415619617](https://doi.org/10.1177/0003122415619617); pmid: [27445413](https://pubmed.ncbi.nlm.nih.gov/27445413/)
89. A. M. O'Rand, J. Hamill-Luker, Processes of cumulative adversity: Childhood disadvantage and increased risk of heart attack across the life course. *J. Gerontol. B Psychol. Sci. Soc. Sci.* **60**, 117–124 (2005). doi: [10.1093/geronb/60.Special_Issue_2.S117](https://doi.org/10.1093/geronb/60.Special_Issue_2.S117); pmid: [16251582](https://pubmed.ncbi.nlm.nih.gov/16251582/)
90. H. C. S. Meier et al., Early life socioeconomic position and immune response to persistent infections among elderly Latinos. *Soc. Sci. Med.* **166**, 77–85 (2016). doi: [10.1016/j.socscimed.2016.07.004](https://doi.org/10.1016/j.socscimed.2016.07.004); pmid: [27543684](https://pubmed.ncbi.nlm.nih.gov/27543684/)
91. M. M. Kittleson et al., Association of childhood socioeconomic status with subsequent coronary heart disease in physicians. *Arch. Intern. Med.* **166**, 2356–2361 (2006). doi: [10.1001/archinte.166.21.2356](https://doi.org/10.1001/archinte.166.21.2356); pmid: [17130389](https://pubmed.ncbi.nlm.nih.gov/17130389/)
92. L. Gaydosch, K. M. Schorpp, E. Chen, G. E. Miller, K. M. Harris, College completion predicts lower depression but higher metabolic syndrome among disadvantaged minorities in young adulthood. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 109–114 (2018). doi: [10.1073/pnas.1714616114](https://doi.org/10.1073/pnas.1714616114); pmid: [29255040](https://pubmed.ncbi.nlm.nih.gov/29255040/)
93. G. E. Miller, T. Yu, E. Chen, G. H. Brody, Self-control forecasts better psychosocial outcomes but faster epigenetic aging in low-SES youth. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 10325–10330 (2015). doi: [10.1073/pnas.1505063112](https://doi.org/10.1073/pnas.1505063112); pmid: [26170291](https://pubmed.ncbi.nlm.nih.gov/26170291/)
94. S. A. James, S. A. Hartnett, W. D. Kalsbeek, John Henryism and blood pressure differences among black men. *J. Behav. Med.* **6**, 259–278 (1983). doi: [10.1007/BF01315113](https://doi.org/10.1007/BF01315113); pmid: [6663614](https://pubmed.ncbi.nlm.nih.gov/6663614/)
95. S. A. James, D. S. Strogatz, S. B. Wing, D. L. Ramsey, Socioeconomic status, John Henryism, and hypertension in blacks and whites. *Am. J. Epidemiol.* **126**, 664–673 (1987). doi: [10.1093/oxfordjournals.aje.a114706](https://doi.org/10.1093/oxfordjournals.aje.a114706); pmid: [3631056](https://pubmed.ncbi.nlm.nih.gov/3631056/)
96. E. B. Cooper, L. E. B. Kruuk, Ageing with a silver-spoon: A meta-analysis of the effect of developmental environment on senescence. *Evol. Lett.* **2**, 460–471 (2018). doi: [10.1002/evl3.79](https://doi.org/10.1002/evl3.79); pmid: [30283695](https://pubmed.ncbi.nlm.nih.gov/30283695/)
97. P. Monaghan, M. F. Haussmann, The positive and negative consequences of stressors during early life. *Early Hum. Dev.* **91**, 643–647 (2015). doi: [10.1016/j.earhumdev.2015.08.008](https://doi.org/10.1016/j.earhumdev.2015.08.008); pmid: [26385447](https://pubmed.ncbi.nlm.nih.gov/26385447/)
98. P. Monaghan, Organismal stress, telomeres and life histories. *J. Exp. Biol.* **217**, 57–66 (2014). doi: [10.1242/jeb.090043](https://doi.org/10.1242/jeb.090043); pmid: [24353204](https://pubmed.ncbi.nlm.nih.gov/24353204/)
99. J. Tung, E. A. Archie, J. Altmann, S. C. Alberts, Cumulative early life adversity predicts longevity in wild baboons. *Nat. Commun.* **7**, 11181 (2016). doi: [10.1038/ncomms11181](https://doi.org/10.1038/ncomms11181); pmid: [27091302](https://pubmed.ncbi.nlm.nih.gov/27091302/)
100. E. D. Strauss, D. Shizuka, K. E. Holekamp, Juvenile rank acquisition is associated with fitness independent of adult rank. *Proc. Biol. Sci.* **287**, 20192969 (2020). doi: [10.1098/rspb.2019.2969](https://doi.org/10.1098/rspb.2019.2969); pmid: [32126950](https://pubmed.ncbi.nlm.nih.gov/32126950/)
101. Y. C. Yang, K. Gerken, K. Schorpp, C. Boen, K. M. Harris, Early-life socioeconomic status and adult physiological functioning: A life course examination of biosocial mechanisms. *Biodemography Soc. Biol.* **63**, 87–103 (2017). doi: [10.1080/19485565.2017.1279536](https://doi.org/10.1080/19485565.2017.1279536); pmid: [28521624](https://pubmed.ncbi.nlm.nih.gov/28521624/)
102. A. E. Willson, K. M. Shuey, G. H. Elder Jr., Cumulative advantage processes as mechanisms of inequality in life course health. *Am. J. Sociol.* **112**, 1886–1924 (2007). doi: [10.1086/512712](https://doi.org/10.1086/512712)
103. K. F. Ferraro, T. P. Shippee, Aging and cumulative inequality: How does inequality get under the skin? *Gerontologist* **49**, 333–343 (2009). doi: [10.1093/geront/gnp034](https://doi.org/10.1093/geront/gnp034); pmid: [19377044](https://pubmed.ncbi.nlm.nih.gov/19377044/)
104. J. Jacobs, K. Agho, G. Stevens, B. Raphael, Do childhood adversities cluster in predictable ways? A systematic review. *Vulnerable Child. Youth Stud.* **7**, 103–115 (2012). doi: [10.1080/17450128.2012.658886](https://doi.org/10.1080/17450128.2012.658886)
105. Y. Ben-Shlomo, D. Kuh, A life course approach to chronic disease epidemiology: Conceptual models, empirical challenges and interdisciplinary perspectives. *Int. J.*

- Epidemiol.* **31**, 285–293 (2002). doi: [10.1093/ije/31.2.285](https://doi.org/10.1093/ije/31.2.285); pmid: [11980781](https://pubmed.ncbi.nlm.nih.gov/11980781/)
106. R. Chetty *et al.*, The fading American dream: Trends in absolute income mobility since 1940. *Science* **356**, 398–406 (2017). doi: [10.1126/science.aal4617](https://doi.org/10.1126/science.aal4617); pmid: [28438988](https://pubmed.ncbi.nlm.nih.gov/28438988/)
 107. M. N. Zippie, E. A. Archie, J. Tung, J. Altmann, S. C. Alberts, Intergenerational effects of early adversity on survival in wild baboons. *eLife* **8**, e47433 (2019). doi: [10.7554/eLife.47433](https://doi.org/10.7554/eLife.47433); pmid: [31549964](https://pubmed.ncbi.nlm.nih.gov/31549964/)
 108. S. Z. Goldenberg, G. Wittemyer, Orphaned female elephant social bonds reflect lack of access to mature adults. *Sci. Rep.* **7**, 14408 (2017). doi: [10.1038/s41598-017-14712-2](https://doi.org/10.1038/s41598-017-14712-2); pmid: [29089603](https://pubmed.ncbi.nlm.nih.gov/29089603/)
 109. T. Pudrovska, B. Anikputa, Early-life socioeconomic status and mortality in later life: An integration of four life-course mechanisms. *J. Gerontol. B Psychol. Sci. Soc. Sci.* **69**, 451–460 (2014). doi: [10.1093/geronb/bgt122](https://doi.org/10.1093/geronb/bgt122); pmid: [24496607](https://pubmed.ncbi.nlm.nih.gov/24496607/)
 110. B. S. McEwen, Understanding the potency of stressful early life experiences on brain and body function. *Metabolism* **57** (Suppl 2), S11–S15 (2008). doi: [10.1016/j.metabol.2008.07.006](https://doi.org/10.1016/j.metabol.2008.07.006); pmid: [18803958](https://pubmed.ncbi.nlm.nih.gov/18803958/)
 111. S. W. C. Chang *et al.*, Neuroethology of primate social behavior. *Proc. Natl. Acad. Sci. U.S.A.* **110** (Suppl 2), 10387–10394 (2013). doi: [10.1073/pnas.1301213110](https://doi.org/10.1073/pnas.1301213110); pmid: [23754410](https://pubmed.ncbi.nlm.nih.gov/23754410/)
 112. R. J. Davidson, B. S. McEwen, Social influences on neuroplasticity: Stress and interventions to promote well-being. *Nat. Neurosci.* **15**, 689–695 (2012). doi: [10.1038/nn.3093](https://doi.org/10.1038/nn.3093); pmid: [22534579](https://pubmed.ncbi.nlm.nih.gov/22534579/)
 113. A. E. Berens, S. K. G. Jensen, C. A. Nelson 3rd, Biological embedding of childhood adversity: From physiological mechanisms to clinical implications. *BMC Med.* **15**, 135 (2017). doi: [10.1186/s12916-017-0895-4](https://doi.org/10.1186/s12916-017-0895-4); pmid: [28724431](https://pubmed.ncbi.nlm.nih.gov/28724431/)
 114. N. Snyder-Mackler, A. J. Lea, Functional genomic insights into the environmental determinants of mammalian fitness. *Curr. Opin. Genet. Dev.* **53**, 105–112 (2018). doi: [10.1016/j.gde.2018.08.001](https://doi.org/10.1016/j.gde.2018.08.001); pmid: [30142491](https://pubmed.ncbi.nlm.nih.gov/30142491/)
 115. N. Adler, N. R. Bush, M. S. Pantell, Rigor, vigor, and the study of health disparities. *Proc. Natl. Acad. Sci. U.S.A.* **109** (suppl. 2), 17154–17159 (2012). doi: [10.1073/pnas.1121399109](https://doi.org/10.1073/pnas.1121399109); pmid: [23045672](https://pubmed.ncbi.nlm.nih.gov/23045672/)
 116. C. A. Shively, T. C. Register, T. B. Clarkson, Social stress, visceral obesity, and coronary artery atherosclerosis: Product of a primate adaptation. *Am. J. Primatol.* **71**, 742–751 (2009). doi: [10.1002/ajp.20706](https://doi.org/10.1002/ajp.20706); pmid: [19452515](https://pubmed.ncbi.nlm.nih.gov/19452515/)
 117. C. A. Shively, Social subordination stress, behavior, and central monoaminergic function in female cynomolgus monkeys. *Biol. Psychiatry* **44**, 882–891 (1998). doi: [10.1016/S0006-3223\(97\)00437-X](https://doi.org/10.1016/S0006-3223(97)00437-X); pmid: [9807643](https://pubmed.ncbi.nlm.nih.gov/9807643/)
 118. J. R. Kaplan, S. B. Manuck, T. B. Clarkson, F. M. Lusso, D. M. Taub, Social status, environment, and atherosclerosis in cynomolgus monkeys. *Arteriosclerosis* **2**, 359–368 (1982). doi: [10.1161/01.ATV.2.5.359](https://doi.org/10.1161/01.ATV.2.5.359); pmid: [6889852](https://pubmed.ncbi.nlm.nih.gov/6889852/)
 119. S. Cohen *et al.*, Chronic social stress, social status, and susceptibility to upper respiratory infections in nonhuman primates. *Psychosom. Med.* **59**, 213–221 (1997). doi: [10.1097/00006842-199705000-00001](https://doi.org/10.1097/00006842-199705000-00001); pmid: [9254393](https://pubmed.ncbi.nlm.nih.gov/9254393/)
 120. G. L. Hermes *et al.*, Social isolation dysregulates endocrine and behavioral stress while increasing malignant burden of spontaneous mammary tumors. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 22393–22398 (2009). doi: [10.1073/pnas.0910753106](https://doi.org/10.1073/pnas.0910753106); pmid: [20018726](https://pubmed.ncbi.nlm.nih.gov/20018726/)
 121. World Health Organization, "World Health Organization, Global Health Estimates 2016: disease burden by cause, age, sex, by country and by region, 2000–2016" (World Health Organization, 2016); www.who.int/healthinfo/global_burden_disease/estimates/en/index1.html.
 122. G. Conti *et al.*, Primate evidence on the late health effects of early-life adversity. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 8866–8871 (2012). doi: [10.1073/pnas.1205340109](https://doi.org/10.1073/pnas.1205340109); pmid: [22615410](https://pubmed.ncbi.nlm.nih.gov/22615410/)
 123. W. R. Swindell, Dietary restriction in rats and mice: A meta-analysis and review of the evidence for genotype-dependent effects on lifespan. *Ageing Res. Rev.* **11**, 254–270 (2012). doi: [10.1016/j.arr.2011.12.006](https://doi.org/10.1016/j.arr.2011.12.006); pmid: [22210149](https://pubmed.ncbi.nlm.nih.gov/22210149/)
 124. E. S. Epel, G. J. Lithgow, Stress biology and aging mechanisms: Toward understanding the deep connection between adaptation to stress and longevity. *J. Gerontol. A Biol. Sci. Med. Sci.* **69** (suppl. 1), S10–S16 (2014). doi: [10.1093/gerona/glu055](https://doi.org/10.1093/gerona/glu055); pmid: [24833580](https://pubmed.ncbi.nlm.nih.gov/24833580/)
 125. B. S. McEwen *et al.*, Mechanisms of stress in the brain. *Nat. Neurosci.* **18**, 1353–1363 (2015). doi: [10.1038/nn.4086](https://doi.org/10.1038/nn.4086); pmid: [26404710](https://pubmed.ncbi.nlm.nih.gov/26404710/)
 126. S. Cohen *et al.*, Chronic stress, glucocorticoid receptor resistance, inflammation, and disease risk. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 5995–5999 (2012). doi: [10.1073/pnas.1118355109](https://doi.org/10.1073/pnas.1118355109); pmid: [22474371](https://pubmed.ncbi.nlm.nih.gov/22474371/)
 127. L. R. Gesquiere *et al.*, Life at the top: Rank and stress in wild male baboons. *Science* **333**, 357–360 (2011). doi: [10.1126/science.1207120](https://doi.org/10.1126/science.1207120); pmid: [21764751](https://pubmed.ncbi.nlm.nih.gov/21764751/)
 128. R. L. Young *et al.*, Conserved transcriptomic profiles underpin monogamy across vertebrates. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 1331–1336 (2019). doi: [10.1073/pnas.1813775116](https://doi.org/10.1073/pnas.1813775116); pmid: [30617061](https://pubmed.ncbi.nlm.nih.gov/30617061/)
 129. C. C. Rittschof *et al.*, Neuromolecular responses to social challenge: Common mechanisms across mouse, stickleback fish, and honey bee. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 17929–17934 (2014). doi: [10.1073/pnas.1420369111](https://doi.org/10.1073/pnas.1420369111); pmid: [25453090](https://pubmed.ncbi.nlm.nih.gov/25453090/)
 130. S. S. Burmeister, E. D. Jarvis, R. D. Fernald, Rapid behavioral and genomic responses to social opportunity. *PLOS Biol.* **3**, e363 (2005). doi: [10.1371/journal.pbio.0030363](https://doi.org/10.1371/journal.pbio.0030363); pmid: [16216088](https://pubmed.ncbi.nlm.nih.gov/16216088/)
 131. N. Snyder-Mackler *et al.*, Social status alters immune regulation and response to infection in macaques. *Science* **354**, 1041–1045 (2016). doi: [10.1126/science.aah3580](https://doi.org/10.1126/science.aah3580); pmid: [27885030](https://pubmed.ncbi.nlm.nih.gov/27885030/)
 132. J. Tung *et al.*, Social environment is associated with gene regulatory variation in the rhesus macaque immune system. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 6490–6495 (2012). doi: [10.1073/pnas.1202734109](https://doi.org/10.1073/pnas.1202734109); pmid: [22493251](https://pubmed.ncbi.nlm.nih.gov/22493251/)
 133. N. Snyder-Mackler *et al.*, Social status alters chromatin accessibility and the gene regulatory response to glucocorticoid stimulation in rhesus macaques. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 1219–1228 (2019). doi: [10.1073/pnas.1811758115](https://doi.org/10.1073/pnas.1811758115); pmid: [30538209](https://pubmed.ncbi.nlm.nih.gov/30538209/)
 134. T. W. McDade *et al.*, Genome-wide analysis of DNA methylation in relation to socioeconomic status during development and early adulthood. *Am. J. Phys. Anthropol.* **169**, 3–11 (2019). doi: [10.1002/ajpa.23800](https://doi.org/10.1002/ajpa.23800); pmid: [30771258](https://pubmed.ncbi.nlm.nih.gov/30771258/)
 135. S. W. Cole, Human social genomics. *PLOS Genet.* **10**, e1004601 (2014). doi: [10.1371/journal.pgen.1004601](https://doi.org/10.1371/journal.pgen.1004601); pmid: [25166010](https://pubmed.ncbi.nlm.nih.gov/25166010/)
 136. A. Danese, B. S. McEwen, Adverse childhood experiences, allostasis, allostatic load, and age-related disease. *Physiol. Behav.* **106**, 29–39 (2012). doi: [10.1016/j.physbeh.2011.08.019](https://doi.org/10.1016/j.physbeh.2011.08.019); pmid: [21888923](https://pubmed.ncbi.nlm.nih.gov/21888923/)
 137. S. W. Cole, The conserved transcriptional response to adversity. *Curr. Opin. Behav. Sci.* **28**, 31–37 (2019). doi: [10.1016/j.cobeha.2019.01.008](https://doi.org/10.1016/j.cobeha.2019.01.008); pmid: [31592179](https://pubmed.ncbi.nlm.nih.gov/31592179/)
 138. J. Sanz *et al.*, Social history and exposure to pathogen signals modulate social status effects on gene regulation in rhesus macaques. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, 20846 (2019). doi: [10.1073/pnas.1820846116](https://doi.org/10.1073/pnas.1820846116); pmid: [31611381](https://pubmed.ncbi.nlm.nih.gov/31611381/)
 139. M. G. Marmot, R. Sapolsky, Of baboons and men: Social circumstances, biology, and the social gradient in health, in *Sociality, Hierarchy, Health: Comparative Biodemography: A Collection of Papers* (National Academies Press, 2014), pp. 365–388.
 140. E. H. M. Sterck, D. P. Watts, C. P. van Schaik, The evolution of female social relationships in nonhuman primates. *Behav. Ecol. Sociobiol.* **41**, 291–309 (1997). doi: [10.1007/s002650050390](https://doi.org/10.1007/s002650050390)
 141. S. Shultz, C. Opie, Q. D. Atkinson, Stepwise evolution of stable sociality in primates. *Nature* **479**, 219–222 (2011). doi: [10.1038/nature10601](https://doi.org/10.1038/nature10601); pmid: [22071768](https://pubmed.ncbi.nlm.nih.gov/22071768/)
 142. P. M. Kappeler, L. Pozzi, Evolutionary transitions toward pair living in nonhuman primates as stepping stones toward more complex societies. *Sci. Adv.* **5**, eaay1276 (2019). doi: [10.1126/sciadv.aay1276](https://doi.org/10.1126/sciadv.aay1276); pmid: [32064318](https://pubmed.ncbi.nlm.nih.gov/32064318/)
 143. R. W. Wrangham, An ecological model of female-bonded primate groups. *Behaviour* **75**, 262–300 (1980). doi: [10.1163/156853980X000447](https://doi.org/10.1163/156853980X000447)
 144. J. B. Silk, Social components of fitness in primate groups. *Science* **317**, 1347–1351 (2007). doi: [10.1126/science.1140734](https://doi.org/10.1126/science.1140734); pmid: [17823344](https://pubmed.ncbi.nlm.nih.gov/17823344/)
 145. S. L. Johnson, L. J. Leedom, L. Muhtadie, The dominance behavioral system and psychopathology: Evidence from self-report, observational, and biological studies. *Psychol. Bull.* **138**, 692–743 (2012). doi: [10.1037/a0027503](https://doi.org/10.1037/a0027503); pmid: [22506751](https://pubmed.ncbi.nlm.nih.gov/22506751/)
 146. M. Zelikowsky *et al.*, The neuropeptide Tac2 controls a distributed brain state induced by chronic social isolation stress. *Cell* **173**, 1265–1279.e19 (2018). doi: [10.1016/j.cell.2018.03.037](https://doi.org/10.1016/j.cell.2018.03.037); pmid: [29775595](https://pubmed.ncbi.nlm.nih.gov/29775595/)
 147. J. L. Hurst *et al.*, Individual recognition in mice mediated by major urinary proteins. *Nature* **414**, 631–634 (2001). doi: [10.1038/414631a](https://doi.org/10.1038/414631a); pmid: [11740558](https://pubmed.ncbi.nlm.nih.gov/11740558/)
 148. G. A. Matthews *et al.*, Dorsal raphe dopamine neurons represent the experience of social isolation. *Cell* **164**, 617–631 (2016). doi: [10.1016/j.cell.2015.12.040](https://doi.org/10.1016/j.cell.2015.12.040); pmid: [26871628](https://pubmed.ncbi.nlm.nih.gov/26871628/)
 149. T. Zhou, C. Sandi, H. Hu, Advances in understanding neural mechanisms of social dominance. *Curr. Opin. Neurobiol.* **49**, 99–107 (2018). doi: [10.1016/j.conb.2018.01.006](https://doi.org/10.1016/j.conb.2018.01.006); pmid: [29428628](https://pubmed.ncbi.nlm.nih.gov/29428628/)
 150. J. P. Capitanio, L. C. Hawley, S. W. Cole, J. T. Cacioppo, A behavioral taxonomy of loneliness in humans and rhesus monkeys (*Macaca mulatta*). *PLOS ONE* **9**, e110307 (2014). doi: [10.1371/journal.pone.0110307](https://doi.org/10.1371/journal.pone.0110307); pmid: [25354040](https://pubmed.ncbi.nlm.nih.gov/25354040/)
 151. D. Nettle, W. E. Frankenhuys, I. J. Rickard, The evolution of predictive adaptive responses in human life history. *Proc. Biol. Sci.* **280**, 20131343 (2013). doi: [10.1098/rspb.2013.1343](https://doi.org/10.1098/rspb.2013.1343); pmid: [23843395](https://pubmed.ncbi.nlm.nih.gov/23843395/)
 152. P. D. Gluckman, M. A. Hanson, H. G. Spencer, Predictive adaptive responses and human evolution. *Trends Ecol. Evol.* **20**, 527–533 (2005). doi: [10.1016/j.tree.2005.08.001](https://doi.org/10.1016/j.tree.2005.08.001); pmid: [16701430](https://pubmed.ncbi.nlm.nih.gov/16701430/)
 153. A. J. Lea, J. Tung, E. A. Archie, S. C. Alberts, Developmental plasticity: Bridging research in evolution and human health. *Evol. Med. Public Health* **2017**, 162–175 (2018). doi: [10.1093/emph/eox019](https://doi.org/10.1093/emph/eox019); pmid: [29424834](https://pubmed.ncbi.nlm.nih.gov/29424834/)
 154. P. D. Gluckman *et al.*, Towards a new developmental synthesis: Adaptive developmental plasticity and human disease. *Lancet* **373**, 1654–1657 (2009). doi: [10.1016/S0140-6736\(09\)60234-8](https://doi.org/10.1016/S0140-6736(09)60234-8); pmid: [19427960](https://pubmed.ncbi.nlm.nih.gov/19427960/)
 155. C. A. Botero, F. J. Weissing, J. Wright, D. R. Rubenstein, Evolutionary tipping points in the capacity to adapt to environmental change. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 184–189 (2015). doi: [10.1073/pnas.1408589111](https://doi.org/10.1073/pnas.1408589111); pmid: [25422451](https://pubmed.ncbi.nlm.nih.gov/25422451/)
 156. C. Laforch, L. Beccara, R. Tollrian, Inducible defenses: The relevance of chemical alarm cues in *Daphnia*. *Limnol. Oceanogr.* **51**, 1466–1472 (2006). doi: [10.4319/lo.2006.51.3.1466](https://doi.org/10.4319/lo.2006.51.3.1466)
 157. T. Uller, S. Nakagawa, S. English, Weak evidence for anticipatory parental effects in plants and animals. *J. Evol. Biol.* **26**, 2161–2170 (2013). doi: [10.1111/jeb.12212](https://doi.org/10.1111/jeb.12212); pmid: [23937440](https://pubmed.ncbi.nlm.nih.gov/23937440/)
 158. A. J. Lea, J. Altmann, S. C. Alberts, J. Tung, Developmental constraints in a wild primate. *Am. Nat.* **185**, 809–821 (2015). doi: [10.1086/681016](https://doi.org/10.1086/681016); pmid: [25996865](https://pubmed.ncbi.nlm.nih.gov/25996865/)
 159. M. Douhard *et al.*, Fitness consequences of environmental conditions at different life stages in a long-lived vertebrate. *Proc. Biol. Sci.* **281**, 20140276 (2014). doi: [10.1098/rspb.2014.0276](https://doi.org/10.1098/rspb.2014.0276); pmid: [24789898](https://pubmed.ncbi.nlm.nih.gov/24789898/)
 160. A. D. Hayward, I. J. Rickard, V. Lummaa, Influence of early-life nutrition on mortality and reproductive success during a subsequent famine in a preindustrial population. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 13886–13891 (2013). doi: [10.1073/pnas.1301817110](https://doi.org/10.1073/pnas.1301817110); pmid: [23918366](https://pubmed.ncbi.nlm.nih.gov/23918366/)
 161. A. D. Hayward, V. Lummaa, Testing the evolutionary basis of the predictive adaptive response hypothesis in a preindustrial human population. *Evol. Med. Public Health* **2013**, 106–117 (2013). doi: [10.1093/emph/eot007](https://doi.org/10.1093/emph/eot007); pmid: [24481192](https://pubmed.ncbi.nlm.nih.gov/24481192/)
 162. A. Berghänel, M. Heistermann, O. Schülke, J. Ostner, Prenatal stress effects in a wild, long-lived primate: Predictive adaptive responses in an unpredictable environment. *Proc. Biol. Sci.* **283**, 20161304 (2016). doi: [10.1098/rspb.2016.1304](https://doi.org/10.1098/rspb.2016.1304); pmid: [27655764](https://pubmed.ncbi.nlm.nih.gov/27655764/)
 163. M. Douhard *et al.*, The influence of weather conditions during gestation on life histories in a wild Arctic ungulate. *Proc. Biol. Sci.* **283**, 20161760 (2016). doi: [10.1098/rspb.2016.1760](https://doi.org/10.1098/rspb.2016.1760); pmid: [27798304](https://pubmed.ncbi.nlm.nih.gov/27798304/)
 164. R. M. Nesse *et al.*, Evolution in health and medicine Sackler colloquium: Making evolutionary biology a basic science for medicine. *Proc. Natl. Acad. Sci. U.S.A.* **107** (suppl. 1), 1800–1807 (2010). doi: [10.1073/pnas.0906224107](https://doi.org/10.1073/pnas.0906224107); pmid: [19918069](https://pubmed.ncbi.nlm.nih.gov/19918069/)
 165. S. C. Stearns, R. M. Nesse, D. R. Govindaraju, P. T. Ellison, Evolution in health and medicine Sackler colloquium: Evolutionary perspectives on health and medicine. *Proc. Natl. Acad. Sci. U.S.A.* **107** (suppl. 1), 1691–1695 (2010). doi: [10.1073/pnas.0914475107](https://doi.org/10.1073/pnas.0914475107); pmid: [20133821](https://pubmed.ncbi.nlm.nih.gov/20133821/)
 166. D. Nettle, M. Bateson, Adaptive developmental plasticity: What is it, how can we recognize it and when can it evolve? *Proc. Biol. Sci.* **282**, 20151005 (2015). doi: [10.1098/rspb.2015.1005](https://doi.org/10.1098/rspb.2015.1005); pmid: [26203000](https://pubmed.ncbi.nlm.nih.gov/26203000/)

167. K. M. West, E. Blacksher, W. Burke, Genomics, health disparities, and missed opportunities for the nation's research agenda. *JAMA* **317**, 1831–1832 (2017). doi: [10.1001/jama.2017.3096](https://doi.org/10.1001/jama.2017.3096); pmid: [28346599](https://pubmed.ncbi.nlm.nih.gov/28346599/)
168. P. Brodin *et al.*, Variation in the human immune system is largely driven by non-heritable influences. *Cell* **160**, 37–47 (2015). doi: [10.1016/j.cell.2014.12.020](https://doi.org/10.1016/j.cell.2014.12.020); pmid: [25594173](https://pubmed.ncbi.nlm.nih.gov/25594173/)
169. J. C. Beehner, T. J. Bergman, The next step for stress research in primates: To identify relationships between glucocorticoid secretion and fitness. *Horm. Behav.* **91**, 68–83 (2017). doi: [10.1016/j.yhbeh.2017.03.003](https://doi.org/10.1016/j.yhbeh.2017.03.003); pmid: [28284709](https://pubmed.ncbi.nlm.nih.gov/28284709/)
170. F. Colchero *et al.*, The emergence of longevous populations. *Proc. Natl. Acad. Sci. U.S.A.* **113**, E7681–E7690 (2016). doi: [10.1073/pnas.1612191113](https://doi.org/10.1073/pnas.1612191113); pmid: [27872299](https://pubmed.ncbi.nlm.nih.gov/27872299/)
171. R. M. Sapolsky, L. J. Share, A pacific culture among wild baboons: Its emergence and transmission. *PLOS Biol.* **2**, E106 (2004). doi: [10.1371/journal.pbio.0020106](https://doi.org/10.1371/journal.pbio.0020106); pmid: [15094808](https://pubmed.ncbi.nlm.nih.gov/15094808/)
172. J. A. Anderson *et al.*, The costs of competition: high social status males experience accelerated epigenetic aging in wild baboons. *bioRxiv* 961052 [Preprint] 24 February 2020. doi: [10.1101/2020.02.22.961052](https://doi.org/10.1101/2020.02.22.961052)
173. D. W. Brown *et al.*, Adverse childhood experiences and the risk of premature mortality. *Am. J. Prev. Med.* **37**, 389–396 (2009). doi: [10.1016/j.amepre.2009.06.021](https://doi.org/10.1016/j.amepre.2009.06.021); pmid: [19840693](https://pubmed.ncbi.nlm.nih.gov/19840693/)
174. National Center for Health Statistics, National Health Interview Study, 2015 (2015); http://ftp.cdc.gov/pub/Health_Statistics/NCHS/NHIS/SHS.
175. S. A. Fritz, O. R. P. Bininda-Emonds, A. Purvis, Geographical variation in predictors of mammalian extinction risk: Big is bad, but only in the tropics. *Ecol. Lett.* **12**, 538–549 (2009). doi: [10.1111/j.1461-0248.2009.01307.x](https://doi.org/10.1111/j.1461-0248.2009.01307.x); pmid: [19392714](https://pubmed.ncbi.nlm.nih.gov/19392714/)
176. S. M. Durant, Factors affecting life and death in Serengeti cheetahs: Environment, age, and sociality. *Behav. Ecol.* **15**, 11–22 (2004). doi: [10.1093/beheco/arg098](https://doi.org/10.1093/beheco/arg098)
177. E. S. Almberg *et al.*, Social living mitigates the costs of a chronic illness in a cooperative carnivore. *Ecol. Lett.* **18**, 660–667 (2015). doi: [10.1111/ele.12444](https://doi.org/10.1111/ele.12444); pmid: [25983011](https://pubmed.ncbi.nlm.nih.gov/25983011/)
178. L. L. Getz, B. McGuire, T. Pizzuto, J. E. Hofmann, B. Frase, Social organization of the prairie vole (*Microtus ochrogaster*). *J. Mammal.* **74**, 44–58 (1993). doi: [10.2307/1381904](https://doi.org/10.2307/1381904)
179. Y. Gager, O. Gimenez, M. T. O'Mara, D. K. N. Dechmann, Group size, survival and surprisingly short lifespan in socially foraging bats. *BMC Ecol.* **16**, 2 (2016). doi: [10.1186/s12898-016-0056-1](https://doi.org/10.1186/s12898-016-0056-1); pmid: [26767616](https://pubmed.ncbi.nlm.nih.gov/26767616/)
180. A. Barocas, A. Ilany, L. Koren, M. Kam, E. Geffen, Variance in centrality within rock hyrax social networks predicts adult longevity. *PLOS ONE* **6**, e22375 (2011). doi: [10.1371/journal.pone.0022375](https://doi.org/10.1371/journal.pone.0022375); pmid: [21818314](https://pubmed.ncbi.nlm.nih.gov/21818314/)
181. J. Lehmann, B. Majolo, R. McFarland, The effects of social network position on the survival of wild Barbary macaques, *Macaca sylvanus*. *Behav. Ecol.* **27**, 20–28 (2016). doi: [10.1093/beheco/arv169](https://doi.org/10.1093/beheco/arv169)
182. J. Rogers *et al.*, The comparative genomics and complex population history of *Papio* baboons. *Sci. Adv.* **5**, eaau6947 (2019). doi: [10.1126/sciadv.aau6947](https://doi.org/10.1126/sciadv.aau6947); pmid: [30854422](https://pubmed.ncbi.nlm.nih.gov/30854422/)
183. A. Panagakis, S. Hamel, S. D. Côté, Influence of early reproductive success on longevity and late reproductive success in an alpine ungulate. *Am. Nat.* **189**, 667–683 (2017). doi: [10.1086/691388](https://doi.org/10.1086/691388); pmid: [28514626](https://pubmed.ncbi.nlm.nih.gov/28514626/)
184. A. Pusey, J. Williams, J. Goodall, The influence of dominance rank on the reproductive success of female chimpanzees. *Science* **277**, 828–831 (1997). doi: [10.1126/science.277.5327.828](https://doi.org/10.1126/science.277.5327.828); pmid: [9242614](https://pubmed.ncbi.nlm.nih.gov/9242614/)
185. L. M. Fedigan, L. Fedigan, S. Gouzoules, H. Gouzoules, N. Koyama, Lifetime reproductive success in female Japanese macaques. *Folia Primatol. (Basel)* **47**, 143–157 (1986). doi: [10.1159/000156271](https://doi.org/10.1159/000156271); pmid: [3583151](https://pubmed.ncbi.nlm.nih.gov/3583151/)
186. C. Packer, D. A. Collins, A. Sindimwo, J. Goodall, Reproductive constraints on aggressive competition in female baboons. *Nature* **373**, 60–63 (1995). doi: [10.1038/373060a0](https://doi.org/10.1038/373060a0); pmid: [7800039](https://pubmed.ncbi.nlm.nih.gov/7800039/)
187. V. Berger, J.-F. Lemaître, D. Allainé, J.-M. Gaillard, A. Cohas, Early and adult social environments shape sex-specific actuarial senescence patterns in a cooperative breeder. *Am. Nat.* **192**, 525–536 (2018). doi: [10.1086/695913](https://doi.org/10.1086/695913); pmid: [30205028](https://pubmed.ncbi.nlm.nih.gov/30205028/)
188. World Health Organization, About social determinants of health (2017); www.who.int/social_determinants/sdh_definition/en.
189. A. Sánchez-Tójar, J. Schroeder, D. R. Farine, A practical guide for inferring reliable dominance hierarchies and estimating their uncertainty. *J. Anim. Ecol.* **87**, 594–608 (2018). doi: [10.1111/1365-2656.12776](https://doi.org/10.1111/1365-2656.12776); pmid: [29083030](https://pubmed.ncbi.nlm.nih.gov/29083030/)
190. P. Braveman, L. Gottlieb, The social determinants of health: It's time to consider the causes of the causes. *Public Health Rep.* **129** (suppl. 2), 19–31 (2014). doi: [10.1177/00333549142915206](https://doi.org/10.1177/00333549142915206); pmid: [24385661](https://pubmed.ncbi.nlm.nih.gov/24385661/)
191. S. E. Taylor, R. L. Repetti, T. Seeman, Health psychology: What is an unhealthy environment and how does it get under the skin? *Annu. Rev. Psychol.* **48**, 411–447 (1997). doi: [10.1146/annurev.psych.48.1.411](https://doi.org/10.1146/annurev.psych.48.1.411); pmid: [9046565](https://pubmed.ncbi.nlm.nih.gov/9046565/)
192. M. Marmot, R. G. Wilkinson, Social organization, stress, and health, in *Social Determinants of Health* (2005), pp. 6–30.
193. J. C. Beehner, T. J. Bergman, D. L. Cheney, R. M. Seyfarth, P. L. Whitten, Testosterone predicts future dominance rank and mating activity among male chacma baboons. *Behav. Ecol. Sociobiol.* **59**, 469–479 (2006). doi: [10.1007/s00265-005-0071-2](https://doi.org/10.1007/s00265-005-0071-2)
194. E. A. Archie, J. Altmann, S. C. Alberts, Social status predicts wound healing in wild baboons. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 9017–9022 (2012). doi: [10.1073/pnas.1206391109](https://doi.org/10.1073/pnas.1206391109); pmid: [22615389](https://pubmed.ncbi.nlm.nih.gov/22615389/)
195. R. McFarland, B. Majolo, Coping with the cold: Predictors of survival in wild Barbary macaques, *Macaca sylvanus*. *Biol. Lett.* **9**, 20130428 (2013). doi: [10.1098/rsbl.2013.0428](https://doi.org/10.1098/rsbl.2013.0428); pmid: [23804292](https://pubmed.ncbi.nlm.nih.gov/23804292/)
196. R. M. Seyfarth, A model of social grooming among adult female monkeys. *J. Theor. Biol.* **65**, 671–698 (1977). doi: [10.1016/0022-5193\(77\)90015-7](https://doi.org/10.1016/0022-5193(77)90015-7); pmid: [406485](https://pubmed.ncbi.nlm.nih.gov/406485/)
197. N. Snyder-Mackler *et al.*, Social status drives social relationships in groups of unrelated female rhesus macaques. *Anim. Behav.* **111**, 307–317 (2016). doi: [10.1016/j.anbehav.2015.10.033](https://doi.org/10.1016/j.anbehav.2015.10.033); pmid: [26769983](https://pubmed.ncbi.nlm.nih.gov/26769983/)
198. O. Schülke, J. Bhagavatula, L. Vigilant, J. Ostner, Social bonds enhance reproductive success in male macaques. *Curr. Biol.* **20**, 2207–2210 (2010). doi: [10.1016/j.cub.2010.10.058](https://doi.org/10.1016/j.cub.2010.10.058); pmid: [21093261](https://pubmed.ncbi.nlm.nih.gov/21093261/)

ACKNOWLEDGMENTS

We thank F. Florey Eischen for her invaluable support in organizing this working group, M. Shanahan and J. Buher Kane for their contributions to early discussions, and three anonymous reviewers for constructive feedback on early versions of this manuscript. **Funding:** The Triangle Area Social and Biological Determinants of Health Working Group is supported by the Duke Center for Population Health and Aging (with funding from NIH P30AG034424), the Carolina Population Center (with funding from NIH P2C HD050924), the Triangle Center for Evolutionary Medicine, and the Duke Social Sciences Research Institute. This work was also supported by NIH grants R01AG057235, R01HD088558, and R01GM102562 to J.T.; R00AG051764, R01AG060931, and T32AG000139 to N.S.-M.; R01HL087103 to C.A.S.; R01AG057800 to Y.C.Y.; R24AG065172 to J.T., A.B., and K.M.H.; R01AG053308 and P01AG031719 to S.C.A.; P01HD031921 and R01HD087061 to K.M.H.; F32HD084117 to L.G.; R01DK102496 to A.B.; R01MD013349, R01MD011728, and T32HD091058 to A.E.A.; a Human Frontier Science Program Research Grant to J.T.; a Minnesota Partnership for Biotechnology and Medical Genomics 18.04 to A.B.; and a Jacobs Foundation Early Career Fellowship to D.W.B.; G.A.N. received support from NIH T32HD091058 and NIH 1K99AG062749-01A1, L.G. received support from NIH T32HD007168, J.R.B. received support from the Bridging Biodiversity and Conservation Science program at the University of Arizona, and G.A.N. and N.S.-M. received support from NIH T32AG000029. **Author contributions:** J.T., N.S.-M., L.G., C.A.S., and S.C.A. wrote the initial draft of the paper; N.S.-M., J.R.B., F.A.C., S.C.A., and J.T. assembled figures; J.R.B., D.W.B., F.A.C., G.A.N., A.B., Y.C.Y., A.E.A., A.O., C.A.S., and K.M.H. contributed edits to subsequent revisions; all authors contributed to the ideas represented in the review. **Competing interests:** Authors declare no competing interests. **Data availability:** All data presented here are based on previously published data sets available in ([2](https://doi.org/10.1016/j.cub.2010.10.058), [8](https://doi.org/10.1016/j.cub.2010.10.058), [31](https://doi.org/10.1016/j.cub.2010.10.058), [37–39](https://doi.org/10.1016/j.cub.2010.10.058)). R code for generating the figures is available at www.tung-lab.org/protocols-and-software.html.

10.1126/science.aax9553

RESEARCH ARTICLE SUMMARY

PLANT SCIENCE

Horizontal gene transfer of *Fhb7* from fungus underlies *Fusarium* head blight resistance in wheat

Hongwei Wang^{*†}, Silong Sun^{*}, Wenyang Ge^{*}, Lanfei Zhao^{*}, Bingqian Hou^{*}, Kai Wang^{*}, Zhongfan Lyu^{*}, Liyang Chen, Shoushen Xu, Jun Guo, Min Li, Peisen Su, Xuefeng Li, Guiping Wang, Cunyao Bo, Xiaojian Fang, Wenwen Zhuang, Xinxin Cheng, Jianwen Wu, Luhao Dong, Wuying Chen, Wen Li, Guilian Xiao, Jinxiao Zhao, Yongchao Hao, Ying Xu, Yu Gao, Wenjing Liu, Yanhe Liu, Huayan Yin, Jiazhu Li, Xiang Li, Yan Zhao, Xiaoqian Wang, Fei Ni, Xin Ma, Anfei Li, Steven S. Xu, Guihua Bai, Eviatar Nevo, Caixia Gao, Herbert Ohm, Lingrang Kong[†]

INTRODUCTION: *Fusarium* head blight (FHB) is a fungal disease that devastates global wheat production, with losses of billions of dollars annually. Unlike foliar diseases, FHB occurs directly on wheat spikes (inflorescences). The infection lowers grain yield and also causes the grain to be contaminated by mycotoxins produced by the *Fusarium* pathogen, thus imposing health threats to humans and livestock. Although plant breeders have improved wheat resistance to FHB, the lack of wheat strains with stable FHB resistance has limited progress.

RATIONALE: Many genetic loci in wheat affect FHB resistance but most only have minor

effects; only a few exhibit a stable major effect on resistance. Wheat relatives in the Triticeae tribe carry resistant genes to different diseases including FHB and thus can be alternative sources of FHB resistance for wheat breeding. *Thinopyrum* wheatgrass has been used as a source of beneficial genes transferable to wheat by distant hybridization breeding since the 1930s. *Fhb7*, a gene transferred from *Thinopyrum* to wheat, showed a stable large effect on FHB resistance. However, the lack of a *Thinopyrum* reference genome hampered gene cloning and marker development, delaying the use of *Fhb7* in wheat breeding. Here, we cloned *Fhb7* using a reference assembly

that we generated for *Th. elongatum* and characterized its resistance mechanisms and evolutionary history.

RESULTS: Using sequence data from *Th. elongatum*, we assembled the Triticeae E reference genome with 44,474 high-confidence genes annotated. Using this reference, we genetically mapped *Fhb7* and located it to a 245-kb genomic region. We determined a gene

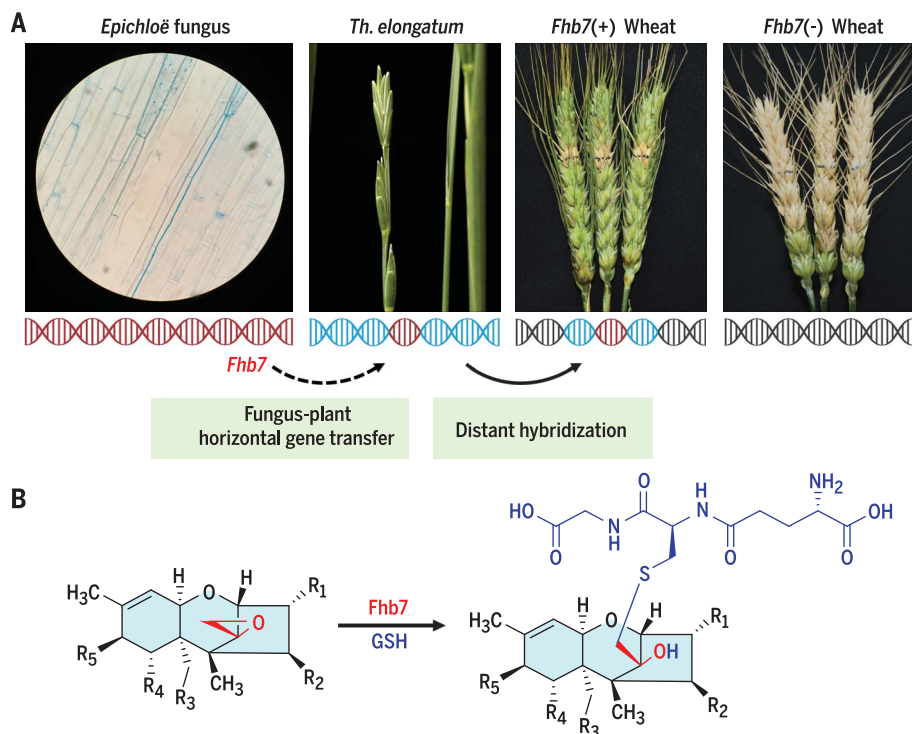
ON OUR WEBSITE

Read the full article at <https://dx.doi.org/10.1126/science.aba5435>

encoding a glutathione S-transferase (GST) as *Fhb7* by virus-induced gene silencing and evaluated mutants and transgenic plants. We discovered that

Fhb7 detoxifies pathogen-produced trichothecene toxins by conjugating a glutathione (GSH) unit onto the epoxide moieties of type A and B trichothecenes. *Fhb7* GST homologs are absent in the plant kingdom, but one sequence showing ~97% identity with *Fhb7* was found in endophytic fungi of an *Epichloë* species that establishes symbiosis with temperate grasses. This result suggests that *Fhb7* might have been transferred from *Epichloë* to *Th. elongatum* through horizontal gene transfer. Finally, we demonstrated that *Fhb7*, when introgressed into diverse wheat backgrounds by distant hybridization, confers broad resistance to both FHB and crown rot without penalizing wheat yield. Our results suggest a source of *Fusarium* resistance for wheat improvement.

CONCLUSION: *Th. elongatum* carries biotic and abiotic resistance genes and is a useful resource for wheat breeding. The assembled *Th. elongatum* reference genome can aid identification and cloning of such genes for wheat improvement. Cloning of *Fhb7* revealed that it encodes a GST that can detoxify trichothecene toxins. Thus, *Fhb7* resistance differs from *Fhb1* resistance, which depends on a reduction of pathogen growth in spikes, although both confer durable resistance. The ability of *Fhb7* to detoxify multiple mycotoxins produced by various *Fusarium* species demonstrates its potential as a source of resistance to the various diseases for which *Fusarium* trichothecenes are virulence factors. The deployment of *Fhb7* in commercial wheat cultivars could alleviate both the food safety issue for consumers and the yield loss problem for growers. Sequence homologies between fungal and plant *Fhb7* suggested that horizontal gene transfer may help to shape plant genomes. ■



***Fhb7* confers FHB resistance by detoxifying trichothecenes.** (A) *Fhb7* in *Th. elongatum* genome likely came from an *Epichloë* fungus through horizontal gene transfer. *Fhb7* drives FHB resistance when introgressed from *Thinopyrum* into wheat. (B) *Fhb7* encodes a GST that detoxifies *Fusarium*-produced trichothecenes by conjugating GSH (blue) to the epoxy group (red). R₁ to R₅ refer to the variable groups in trichothecenes.

The list of author affiliations is available in the full article online.

*These authors contributed equally to this work.

†Corresponding author. Email: lkong@sdau.edu.cn (L.K.); wanghongwei@sdau.edu.cn (H.W.)

Cite this article as H. Wang et al., *Science* 368, eaba5435 (2020). DOI: 10.1126/science.aba5435

RESEARCH ARTICLE

PLANT SCIENCE

Horizontal gene transfer of *Fhb7* from fungus underlies *Fusarium* head blight resistance in wheat

Hongwei Wang^{1*†}, Silong Sun^{1*}, Wenyang Ge^{1*}, Lanfei Zhao^{1*}, Bingqian Hou^{1*}, Kai Wang^{2*}, Zhongfan Lyu^{1*}, Liyang Chen², Shoushen Xu¹, Jun Guo³, Min Li¹, Peisen Su¹, Xuefeng Li¹, Guiping Wang¹, Cunyao Bo¹, Xiaojian Fang¹, Wenwen Zhuang¹, Xinxin Cheng¹, Jianwen Wu¹, Luhao Dong¹, Wuying Chen¹, Wen Li¹, Guilian Xiao¹, Jinxiao Zhao¹, Yongchao Hao¹, Ying Xu¹, Yu Gao¹, Wenjing Liu¹, Yanhe Liu¹, Huayan Yin¹, Jiazhu Li⁴, Xiang Li¹, Yan Zhao¹, Xiaoqian Wang¹, Fei Ni¹, Xin Ma¹, Anfei Li¹, Steven S. Xu⁵, Guihua Bai⁶, Eviatar Nevo⁷, Caixia Gao⁸, Herbert Ohm⁹, Lingrang Kong^{1†}

Fusarium head blight (FHB), a fungal disease caused by *Fusarium* species that produce food toxins, currently devastates wheat production worldwide, yet few resistance resources have been discovered in wheat germplasm. Here, we cloned the FHB resistance gene *Fhb7* by assembling the genome of *Thinopyrum elongatum*, a species used in wheat distant hybridization breeding. *Fhb7* encodes a glutathione S-transferase (GST) and confers broad resistance to *Fusarium* species by detoxifying trichothecenes through de-epoxidation. *Fhb7* GST homologs are absent in plants, and our evidence supports that *Th. elongatum* has gained *Fhb7* through horizontal gene transfer (HGT) from an endophytic *Epichloë* species. *Fhb7* introgressions in wheat confers resistance to both FHB and crown rot in diverse wheat backgrounds without yield penalty, providing a solution for *Fusarium* resistance breeding.

Wheat (*Triticum aestivum* L.) is a leading source of calories for the human population (1). The prevalence and widespread outbreaks of the devastating *Fusarium* head blight (FHB) disease, exacerbated by recent changes in climate and certain cropping practices, has posed a threat for global wheat production and food safety. *Fusarium* species cause not only FHB in wheat, barley, and oat, but also crown rot in wheat and ear rot in maize. However, *F. graminearum* is the prominent pathogen of wheat FHB in China, the United States, Canada, Europe, and many other countries (2). *Fusarium* produces epoxy-sesquiterpenoid compounds known as trichothecenes, some examples of which are deoxynivalenol (DON), T-2 toxin, HT-2 toxin, and nivalenol (NIV), among others; these compounds are inhib-

itors of protein synthesis and virulence factors for pathogenicity (2). Trichothecene contamination in cereal grain results in immunotoxicity and cytotoxicity in humans and animals and thus has aroused public safety concerns (3). Despite global screening efforts examining tens of thousands of wheat accessions, a limited number of quantitative trait loci (QTLs) have been verified to confer a stable effect on FHB resistance (4). *Fhb1* on chromosome 3B is the only QTL that has been used in breeding programs worldwide. Although it has been cloned from different Chinese wheat sources, its molecular identity and resistance mechanisms remain equivocal (5–8).

Wheat relatives have proven to be alternative sources for improvement of resistance to both biotic and abiotic stresses in wheat (9). Distant hybridization, the practice of making crosses between two different species, genera, or higher-ranking taxa, makes it possible to transfer alien genes from Triticeae tribe relatives to wheat (9–11). Tall and intermediate wheatgrasses of the *Thinopyrum* genus (forage grasses) are sources of resistance to salinity, drought, and disease for wheat. Several disease resistance genes, including stem rust (e.g., *Sr24*, *Sr25*, *Sr26*, *Sr43*, *Sr44*, and *SrB*), leaf rust (*Lr19*, *Lr24*, *Lr29*, and *Lr38*), powdery mildew (*Pm40* and *Pm43*), barley yellow dwarf virus (*Bdv2* and *Bdv3*), and *Fusarium* head blight (*Fhb7*), have been introduced from *Thinopyrum* into wheat for resistance breeding (10, 12–16).

Fhb7 is a QTL introduced from *Thinopyrum elongatum* and shows a similar effect on FHB re-

sistance as *Fhb1*. *Th. elongatum* (syn. *Agropyron elongatum* or *Lophopyrum elongatum*), a grass of the Triticeae family with a diploid E genome ($2n = 2x = 14$), is native to Eurasia and is thought to be a genome donor species for various tetra-, hexa-, and even decaploid species in the *Thinopyrum* genus (14). The lack of a reference sequence for the E genome has impeded the process of cloning and the development of diagnostic markers for the deployment of *Fhb7* and other E genome-derived resistance genes. Here, we report the assembly of a reference genome for *Th. elongatum* and describe the cloning and biomolecular characterization of *Fhb7*. Using the newly assembled E genome reference, we identified a GST gene as a candidate for *Fhb7* by map-based cloning and confirmed its function in FHB resistance using transgenics. *Fhb7* can detoxify trichothecenes by catalyzing the conjugation of a glutathione (GSH) unit onto their toxic epoxide moiety. *Fhb7*'s coding sequence has no obvious homology to any known sequence in the entire plant kingdom but shares 97% sequence identity with a species of endophytic fungus (*Epichloë aotearoae*) known to infect temperate grasses, which provides evidence that *Fhb7* in the *Th. elongatum* genome might be derived from the fungus through HGT. We demonstrate here that *Fhb7* confers resistance to both FHB and crown rot without yield penalty in wheat.

Results

Th. elongatum genome assembly and evolution

To sequence and assemble the genome of *Th. elongatum*, 1.1 Tb of high-quality sequence reads were generated from a series of libraries, which is about 236× coverage of the *Th. elongatum* genome (table S1). We initially assembled the short sequence reads using DeNovoMAGICTM3.0 software (NRGene) and then filled the gaps using ~145 Gb (~31×) PacBio SMRT reads. The initial assembly was finely tuned using high-quality paired-end polymerase chain reaction (PCR)-free reads. Two Bionano optical maps (based on enzymes BspQI and DLE1 data) were further used to extend the scaffolds (tables S2 and S3), which resulted in a 4.63-Gb assembly with a contig N50 size of 2.15 Mb and a scaffold N50 size of 73.24 Mb (Table 1).

To construct the pseudochromosomes, high-throughput chromosome conformation capture (Hi-C) data were used to categorize and order the assembled scaffolds (table S4). A total of 141 scaffolds were anchored and oriented onto seven pseudochromosomes, which account for 95% of the estimated genome size (4.78 Gb; fig. S1) and 98% of the assembled genome sequences (fig. S2). About 97.6% complete and 1.3% fragmented Embryophyta genes were detected in our assembly according to BUSCO [Benchmarking Universal Single-Copy Orthologs (17)], proportions comparable

¹State Key Laboratory of Crop Biology, College of Agronomy, Shandong Agricultural University, Tai'an, Shandong 271018, PR China. ²Novogene Bioinformatics Institute, Beijing 100083, PR China. ³Crop Research Institute, Shandong Academy of Agricultural Sciences, Jinan, Shandong 250100, PR China. ⁴College of Chemistry and Chemical Engineering, Yantai University, Yantai, Shandong 264005, PR China.

⁵USDA-ARS, Cereal Crops Research Unit, Edward T. Schafer Agricultural Research Center, Fargo, ND 58102, USA.

⁶USDA-ARS, Hard Winter Wheat Genetics Research Unit, Manhattan, KS 66506, USA. ⁷Institute of Evolution, University of Haifa, Mount Carmel, Haifa 3498338, Israel.

⁸State Key Laboratory of Plant Cell and Chromosome Engineering, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, PR China.

⁹Department of Agronomy, Purdue University, West Lafayette, IN 47907, USA.

*These authors contributed equally to this work.

†Corresponding author. Email: lkong@sdaa.edu.cn (L.K.); wanghongwei@sdaa.edu.cn (H.W.)

Table 1. Summary statistics for <i>Th. elongatum</i> genome assembly.	
Assembly characteristics	Values
Estimated genome size	4.78 Gb
Total length of contigs	4.58 Gb
N50 length of contigs	2.15 Mb
Total number of contigs	12,262
Longest contigs	11.6 Mb
Total length of scaffolds	4.63 Gb
N50 length of scaffolds	73.24 Mb
Total number of scaffolds	783
Longest scaffolds	258.71 Mb
Total gap size	52.78 Mb
Total sequences anchored to the pseudochromosomes	4.54 Gb
Number of annotated high-confidence genes	44,474
Percentage of repeat sequences	81.29%
Complete BUSCOs	97.6%
Fragmented BUSCOs	1.3%
Missed BUSCOs	1.1%

to other *Triticum* genomes (table S5). The quality of the E genome assembly was validated by assessment of the long terminal repeat (LTR) completeness using LTR Assembly Index (LAI) software (18) (table S6), by genomic alignment with 61 randomly selected bacterial artificial chromosome (BAC) clones (fig. S3 and table S7), and by the consistency of our assembly with a high-density genetic map from a hexaploid *Thinopyrum* species (19) (fig. S4).

Repetitive elements are dispersed throughout the E genome, with ~81.29% of the *Th. elongatum* assembly being annotated as repetitive elements, including retrotransposons (62.39%), DNA transposons (17.83%), and unclassified elements (1.07%) (table S8 and table S9). Analysis of the Cereba and Quinta LTR retrotransposons supported that the centromere regions were appropriately assembled (fig. S5). The composition of different classes of repetitive DNA in the E genome was similar to those of the wheat A, B, or D subgenomes (fig. S6). No recent LTR burst was detected in the E or common wheat genomes (fig. S7), suggesting relatively stable genomes and helping to explain the success of distant hybridization breeding efforts using these materials. A total of 44,474 high-confidence protein-coding genes were predicted on the basis of a combination of methods [ab initio, protein homology based, and RNA-sequencing (RNA-seq) based], and 44,144 (99.3%) of the predicted genes were anchored onto the seven assembled pseudochromosomes (figs. S8 and S9 and tables S10 to S12).

Gene family analysis identified 32,048 orthologous genes between the E genome and the wheat A, B, or D genome or the barley genome (fig. S10). A synonymous substitution rate (K_s) value was calculated using a moving-

average model with the ortholog dataset, which revealed similar K_s peak values between the E genome and the wheat subgenomes (E and A: 0.0645, E and B: 0.0645, E and D: 0.062), indicating a branching time for *Th. elongatum* and *Triticum* of ~4.77 to 4.96 million years ago when a nucleotide substitution rate of 6.5×10^{-9} was used (Fig. 1A) (20).

We also compared the E genome with other Triticeae genomes that have been used for distant hybridization based on a maximum likelihood tree built using single-copy genes from available Triticeae genome assemblies; the tree also incorporated transcript data for several diploid species, including the Triticeae R, Q, V, F, and Ns genomes (table S13). The three wheat subgenomes are more closely related to the E genome of *Th. elongatum* than they are to the R genome of rye, another species frequently used in wheat distant hybridization (Fig. 1A). A syntenic block analysis indicated genome-wide colinearity between the E genome and the A, B, or D genomes, which helps to explain the success of E-genome-based distant hybridization breeding in wheat (Fig. 1B and data S1). Substantial colinearity notwithstanding, we did identify 18 fragmental inversions between the E genome and the wheat subgenomes, with sizes ranging from 1.5 to 18 Mb, which is supported by both the Bionano maps and Hi-C data (fig. S11 and table S14).

Map-based cloning of the *Fusarium* resistance gene *Fhb7*

A total of 1897 resistance gene analogs (RGA) were annotated in the E genome (fig. S12 and table S15). An apparent RGA expansion, especially for CC-NBS-LRR (CNL), on the distal end of the long arm of chromosome 7E (7EL) is accompanied with the expansion of this

genomic region (fig. S13 and table S16). Some of the alien resistance gene introgressions into wheat are located in this region, including *Lr19*, *Sr25*, *Bdv3*, and *Fhb7* (10, 13, 14).

Previously, we mapped the *Fhb7* to the distal end of the 7EL (based on recombination between 7E1 and 7E2 in a common wheat background) using a recombinant inbred line (RIL) population from a cross between an FHB-susceptible substitution line (7E1/7D) and an FHB-resistant substitution line (7E2/7D) (13, 21). For further mapping of this gene, we developed a segregation population derived from BC₆F₁ with the same cross, in which FHB resistance was tracked as monogenic trait for validation of phenotypes. We also developed a population to promote 7E recombination by introducing the CS *ph1bph1b* locus (fig. S14). Because of the semidominant nature of *Fhb7*, the homozygous offspring of the recombinants were evaluated for FHB resistance. With analysis of 258 recombinants (between the *XBEA5653* and *XsdauK67* markers) screened from 19,200 progeny of BC₆F₁ population, we confirmed that *Fhb7* is positioned between the *XSdauK79* and *XSdauK80* markers within an ~1.2-Mb region based on the E reference genome (Fig. 1C and fig. S15).

Analysis of the RNA-seq data of E reference genome from *Th. elongatum* spikes identified eight expressed genes in the *Fhb7* region (Fig. 1C and table S17). However, when conducting transcriptomics analysis of the parental lines of 7E1/7D (S) and 7E2/7D (R), we found that only two candidate genes (Tel7E01T1020600.1 and Tel7E01T1021800.1) were expressed in a manner specific to the 7E2 genome (the resistant parent) and E reference genome [which also confers FHB resistance (12, 22)] (Fig. 1C and tables S18 and S19). BAC clones containing Tel7E01T1020600.1 were identified from the resistant donor line and new markers (*XsdauK86* and *XsdauK87*) derived from the BAC ends were developed to screen recombinants among 5760 progeny of the segregation population harboring the CS *ph1bph1b* locus (Fig. 1C, fig. S14, and table S20). Analysis of phenotypic data of the three key recombinants verified that *Fhb7* is located between the *XsdauK86* and *XsdauK88* markers, thereby delineating this locus to a 245-kb region containing a single expressed gene: Tel7E01T1020600.1 (Fig. 1C). This gene is present in the E reference genome and 7E2 genome but absent in the susceptible 7E1 genome based on analysis of genomics and transcriptomics data (table S19 and table S21).

Gene expression analysis using quantitative PCR indicated that Tel7E01T1020600.1 was constitutively expressed in all tissues examined, including root, leaf, shoot, and spike (fig. S16). Moreover, barley stripe mosaic virus (BSMV)-induced gene silencing of Tel7E01T1020600.1 in wheat leaves revealed that it conferred

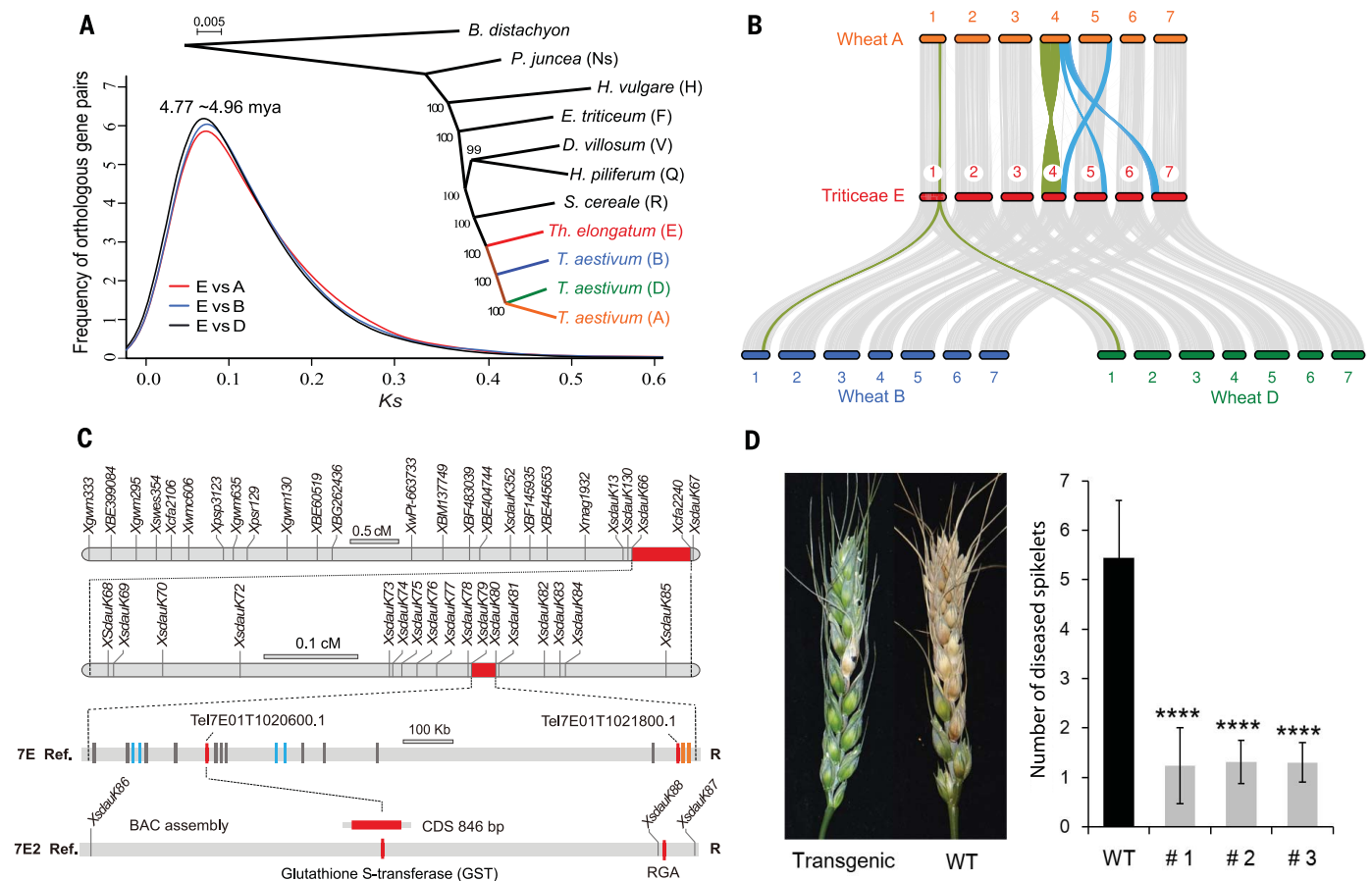


Fig. 1. Genome evolution of *Th. elongatum* and cloning of *Fhb7*. (A) Maximum likelihood phylogenetic tree of the genomes of Triticeae species and the Ks distributions of ortholog genes between the E genome and the wheat Chinese Spring A, B, and D subgenomes. mya, million years ago. (B) Syntenic blocks between the E genome and the three wheat subgenomes. The representative inversion fragment is indicated in green; chromosomal translocations for the wheat A subgenome compared with the E genome are also indicated in blue. (C) Map-based cloning of *Fhb7* at the distal region of chromosome 7E. Using the BC₆F₁ population derived from the cross between two wheat-*Thinopyrum* substitution lines, 7E1/7D and 7E2/7D, *Fhb7* was initially mapped to an interval between the markers *Xsdauk79* and *Xsdauk80* (~1.2 Mb on the E reference genome) (second bar from the top). The expressed genes are labeled as follows: gray refers to no expression in the E reference genome; blue refers to E reference genome-specific expression;

orange refers to expression in the E reference, 7E1 and 7E2 genomes; red refers to expression in FHB-resistant donor genomes of 7E2 and E reference (third bar from the top). BAC clones containing Tel7E01T1020600.1 were identified from the substitution line 7E2/7D, based on which genetic markers (XsdauK86 and XsdauK87) were developed for recombinant screening of the CS *ph1bph1b* population. Finally, *Fhb7* was genetically confirmed within a 245-kb region between markers *XsdauK86* and *XsdauK88*, with only the candidate gene Tel7E01T1020600.1 encoding a GST [CDS is shown in red; untranslated region is shown in gray] (fourth bar from the top). **(D)** FHB was evaluated for wild-type (WT, KN199) and transgenic wheat KN199 expressing the native promoter and the 846-bp open reading frame of *Fhb7*. T₃ plants containing *Fhb7* from three different lines were evaluated for FHB resistance using single floret inoculation (35). The FHB was scored for at least five spikes per repeat, with at least three repeats for each transgenic line.

resistance to *F. graminearum*, supporting that this gene represents *Fhb7* (fig. S17). Sequence analysis of 22 ethyl methanesulfonate (EMS) induced mutants identified five amino acids that were implicated in *Fhb7*'s FHB resistance-related function: S34F, T48I, A98V, A9V, and P106L (fig. S18 and data S2). Moreover, two stop-gain mutations at position 209 or 243 led to reduced resistance to *F. graminearum* (fig. S18 and data S2). To confirm Tel7E01T1020600.1 as *Fhb7*, we transgenically introduced a construct with the native promoter and the 846-base pair (bp) coding sequence of this gene into the FHB-disease-susceptible wheat cultivar KN199 and assessed three indepen-

dent T₃-transgenic plants. The *Fusarium*-inoculated transgenic plants exhibited lower FHB symptom with substantially fewer diseased spikelets per spike than the control (Fig. 1D).

Evolutionary history and molecular function of *Fhb7*

Protein domain-based functional annotation predicted that *Fhb7* likely encodes a GST enzyme. A BLAST search of the *Fhb7* sequence against the National Center for Biotechnology Information (NCBI) GenBank database (23) did not find any homolog of *Fhb7* in the *Triticum* genus or in the entire plant kingdom. How-

ever, there is a homolog sharing 97% identity in the genome of *E. aotearoae* (Fig. 2A and fig. S19). A phylogenetic analysis of the *Fhb7* sequence revealed its distribution among *Epichloë* species, endophytic fungi of temperate grasses (Fig. 2A). Thus, the occurrence of the *Fhb7* gene in the *Th. elongatum* genome might be caused by fungus-to-plant HGT (FP-HGT) event. Because the *Fhb7* locus is present both in the diploid E genome of *Th. elongatum* and in 7E2 from decaploid *Th. ponticum*, this FP-HGT event apparently occurred after the divergence of the E genome from *Triticum* sp. but before the formation of the decaploid *Th. ponticum* (Fig. 2A).

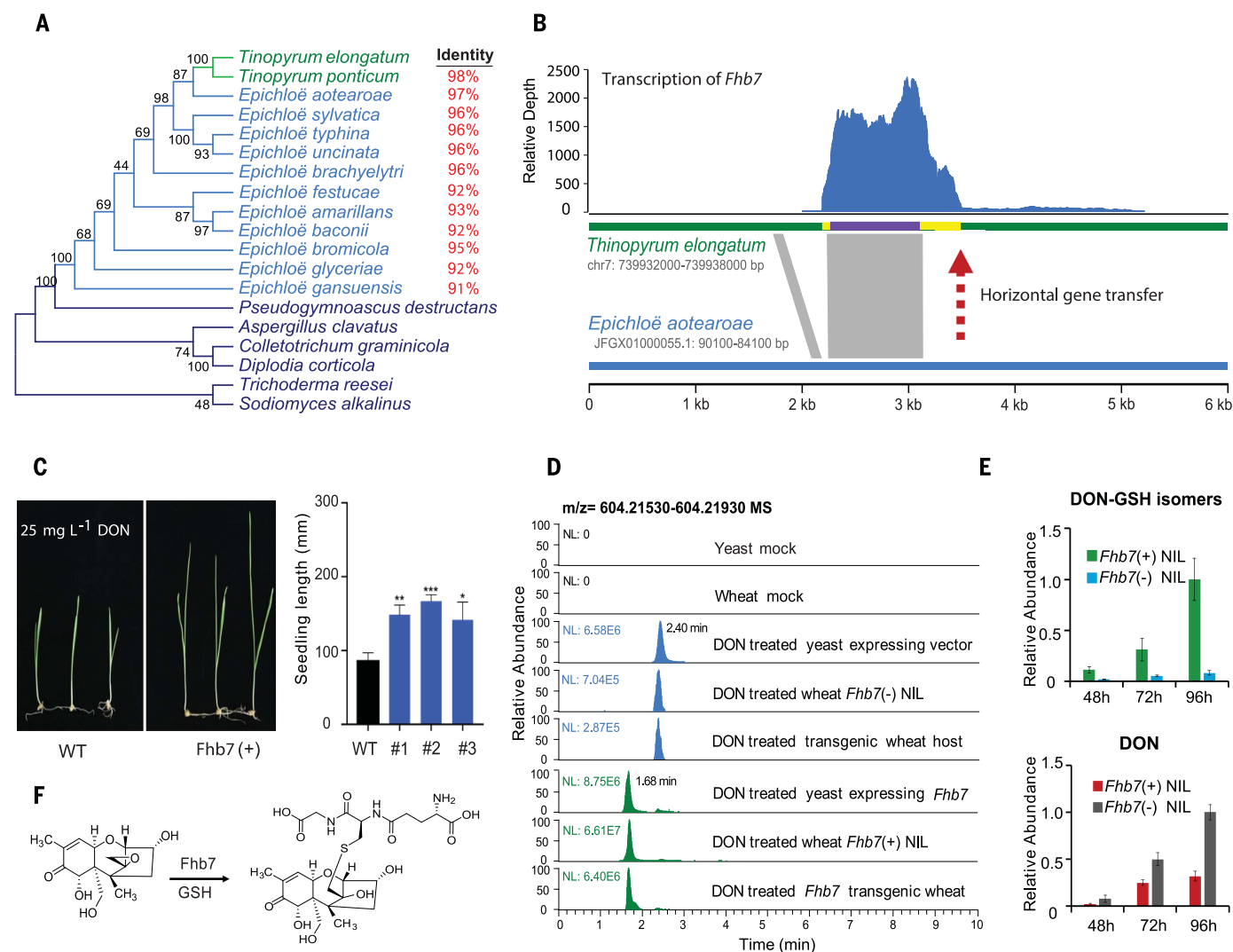


Fig. 2. *Fhb7* confers FHB resistance by detoxifying DON. (A) Maximum likelihood phylogenetic tree of the closest homologs of *Fhb7* from plants and fungi. The DNA sequence similarity with *Fhb7* is marked in red. (B) Horizontal gene transfer of *Fhb7*. The transcripts CDS (purple), and possible untranslated regions (yellow) of *Fhb7* are shown along chromosome 7E, and the sequence sharing high similarity with the *E. aotearoae* genome is presented as a gray block. The genomic fragment (897 bp) containing full CDS and partial untranslated region of *Fhb7* showed 97% identity between the two genomes. (C) DON tolerance of *Fhb7*-transgenic wheat. Seedlings (4 days old) were moved to a petri dish containing 25 mg L⁻¹ DON and seedling length was evaluated 7 d after the DON treatment at room temperature. (D) Extracted ion chromatograms (EICs) at *m/z* 604.2173 revealing the presence of two DON-glutathione adducts. The *Fhb7* NIL, *Fhb7*-transgenic wheat, and

Fhb7-transgenic yeast (*P. pastoris*) cultures were treated with 25 mg L⁻¹ DON for 24 hours. A product that elutes at 1.68 min accumulated in *Fhb7*(+) samples and a known, nonenzymatically produced DON-glutathione adduct product that elutes at 2.4 min accumulated in the corresponding *Fhb7*(-) control samples. (E) Relative abundances of the de-epoxidated *Fhb7*-catalyzed DON-glutathione (green) adduct and the known nonenzymatic DON-glutathione adduct (blue) in spikes of *Fusarium*-challenged NIL plants contrasting in *Fhb7*. After inoculation of *F. graminearum* on spike glumes, the *Fhb7*(+) NIL accumulated a copious amount of de-epoxidated DON-glutathione adduct. By contrast, the DON substrate reduced the accumulation in *Fhb7*(+) NIL compared with that in *Fhb7*(-) NIL, as shown in the bottom bar chart. (F) Molecular structure of the de-epoxidated DON-glutathione adduct catalyzed by *Fhb7*.

The horizontal transfer of the *Fhb7* sequence did not occur as a part of a gene cluster (presuming that it is from *E. aotearoae* as the donor genome; this is the species harboring the closest identified homolog of *Fhb7*) (fig. S20). On the basis of sequence similarity, the sequence was transferred into the diploid E genome as a short fragment, including the 846-bp coding sequence for *Fhb7*, a 32-bp sequence before the start codon, and a 19-bp sequence after the stop codon (Fig. 2B). At the

position 535 bp upstream of *Fhb7*'s start codon in the E genome, another 90-bp sequence shows high identity to a sequence in *E. aotearoae* (Fig. 2B), suggesting the possibility that a larger sequence was initially transferred to *Th. elongatum* but late mutations occurred in the transferred sequence. The insertion of the *Epichloë* genome fragment in the E genome was also identified in a BAC clone harboring *Fhb7* (Fig. 1C and data S3), confirming that the sequence is not an artifact from the genome assembly process.

Phylogenetic analysis of the GST superfamily showed that *Fhb7* belongs to the fungal GTE (glutathione transferase etherase-related) subfamily (fig. S21 and tables S22 and S23), wherein all members contain a LigE domain, but none of which has been functionally characterized to date (24). The *Fhb7* gene is conserved in *Epichloë* species and in multiple *Thinopyrum* species, emphasizing its role in protecting organisms from the cytotoxic damage caused by *Fusarium* species (Fig. 2A and fig. S20).

Gene expression analysis in a time course of *Fusarium* infection in *Th. elongatum* and the 7E2/7D substitution line (table S18) showed that the transcription levels of *Fhb7* were induced at 48 hours after infection (fig. S22).

Research in plant pathology about the progression of *F. graminearum* infection in wheat has established that the fungus starts to produce its DON mycotoxin, an inhibitor of protein synthesis that targets ribosomal machinery, by the 48-hour infection time point (25). We therefore conducted DON assays on wheat seedlings of the 7E2/7D substitution line. The results showed that the expression of *Fhb7* can be induced within 6 hours after DON treatment (fig. S22), suggesting that this putative GST enzyme may have a role in xenobiotic detoxification. To test this hypothesis, we conducted a growth inhibition assay by growing *Fhb7* near-isogenic lines (NILs) and *Fhb7*-transgenic wheat seedlings in media containing DON and found that the plants with *Fhb7* grew better (assessed as seedling length) than the plants without *Fhb7* (Fig. 2C and fig. S23). We also expressed *Fhb7* in yeast to test its growth on DON-containing media and found that both the *Fhb7*(+) and *Fhb7*(-) yeasts grew well in the absence of DON; however, only the *Fhb7*(+) yeast grew normally on the media containing 400 mg L⁻¹ DON (fig. S24).

Further evidence for the involvement of *Fhb7* in detoxification was demonstrated by its direct use of DON as a substrate. We treated the seedlings of NILs, *Fhb7*-transgenic wheat, and *Fhb7*-expressing yeast cultures with DON, and found that the presence of *Fhb7* in wheat and yeast caused accumulation of a chromatographic peak at 1.68 min, but the accumulation was not detected in the corresponding control samples without *Fhb7* (Fig. 2D). This peak had a mass/charge (*m/z*) value of 604.2173 (± 3 ppm) under positive ion mode, which is equal to the value for the molecule comprising DON (296.1259), a glutathione group (307.0838), and a hydrogen atom (1.0078), therefore suggesting that *Fhb7* confers GST activity to form a glutathione adduct of DON (DON-GSH) (Fig. 2D and fig. S25).

Previous studies on FHB- and DON-associated chemistry (26, 27) using nuclear magnetic resonance spectroscopy confirmed the nonenzymatic formation of a DON-GSH adduct that was formed through a reaction with the double bond at C10 on DON's first planar ring. This product was mainly detected in the DON-treated *Fhb7*(-) yeast cultures and *Fhb7*(-) wheat samples with the peak at 2.4 min (Fig. 2D and fig. S25). Although the two detected DON-GSH isomers had identical *m/z* values, tandem mass spectrometry with collision-induced dissociation experiments unequivocally supported that the *Fhb7*(+) samples produce a de-epoxidated DON-GSH adduct (figs. S25 to S28); that is, the GSH group added by *Fhb7*

is attached to the C13 carbon, which disrupts the epoxy group known to be critical in DON's toxicity (Fig. 2F) (28). Further, we used liquid chromatography-high-resolution mass spectrometry (LC-HRMS) to profile DON-treated spikes from 37 diverse wheat germplasm accessions and cultivars without *Fhb7*. We detected the DON-GSH (C10) peak at 2.4 min in all of these plants but did not detect the 1.68-min de-epoxidated DON-GSH (C13) adduct in any of them (fig. S29).

Fusarium species produce a series of trichothecene mycotoxins, including DON, 3-ADON, 15-ADON, T-2, HT-2, fusarenon-X, NIV, diacetoxyscirpenol, and others, the distribution of which varies among *Fusarium* chemotypes (24, 26). Considering the common occurrence of epoxy groups in these trichothecene compounds, we hypothesized that *Fhb7* may be able to detoxify trichothecenes other than DON. Indeed, LC-HRMS analysis of trichothecene-treated wheat samples revealed the presence of GSH adducts for all the trichothecenes that we tested in this study (figs. S30 to S37). In light of *Fhb7*'s wide catalytic spectrum for these mycotoxins, we investigated whether it can confer resistance to other *Fusarium* chemotypes, including *F. pseudograminearum* for crown rot and *F. asiaticum*, a predominant FHB-causing strain in south China. Assays using detached wheat leaves showed that the *Fhb7*-transgenic plants exhibited smaller lesions than wild-type plants for all the tested *Fusarium* species (fig. S38). *F. pseudograminearum* was also inoculated on the base of wheat seedlings, and the results confirmed that the transgenic plants also exhibit improved crown rot resistance compared with the nontransgenic controls (fig. S39). These results further demonstrate how *Th. elongatum* benefits from *Fhb7* through the FP-HGT event, which protects plants from *Fusarium*-caused cytotoxic damage by detoxifying trichothecene through de-epoxidation (fig. S20).

Application of *Fhb7* in *Fusarium* resistance breeding

Considering *Fhb7*'s functionality, specifically in the enzymatic conversion of trichothecenes, we speculated that incorporating the *Fhb7* locus into wheat may confer resistance in different genetic backgrounds without affecting yield traits. Indeed, the translocation of a short fragment [with ~16% of the 7E long arm (13)] on wheat 7D resulted in wheat lines with broad resistance to both FHB and crown rot (Fig. 3, A to C). Detailed characterization of NILs (LX99 background) in field conditions showed no significant difference in agronomic yield traits (e.g., thousand grain weight, flag leaf length, etc.; Fig. 3, D and E). Obvious yield penalty caused by *Fhb7* resistance was also not detected when it was transferred into seven additional genetic backgrounds (Fig. 3F and fig. S40).

These results demonstrated the advantages of *Fhb7*-mediated resistance over other QTLs, including high resistance to both FHB and crown rot and detoxifying DON without yield penalty, and thus highlighted the potential utility of the *Fhb7* locus in future wheat breeding for improved FHB resistance and good yield traits.

Discussion

Fusarium diseases are economically impactful because of their effects on the production of cereal crops. In this study, the successful cloning of *Fhb7* from the Triticeae E genome and characterization of its molecular mechanism advances the knowledge on the essential role of trichothecenes in the pathogenesis of *Fusarium*. We have demonstrated that *Fhb7* confers FHB resistance in diverse wheat genetic backgrounds without yield penalty and *Fhb7* is able to biochemically detoxify trichothecene mycotoxins produced by multiple *Fusarium* species, which highlights the value of *Fhb7* in combating FHB and reducing DON contamination in wheat and other cereal crops through breeding.

The epoxides at the C12/13 of trichothecene mycotoxins are the key contributors to their toxicity. However, to date, genes or proteins with de-epoxidation function have not been identified (3). *Fusarium* species can reduce DON toxicity by adding an acetyl group on the hydroxyl group at C3 and C15, forming 3-ADON and 15-ADON, respectively; however, the reduction of cytotoxicity for these DON derivatives is modest in plant cells (3). In planta, glucosylation at C3 has been documented to detoxify DON by forming DON-3-glucoside (D3G), which is reversible in animals, causing release of DON during digestion (29). Here, beyond our identification of an FHB resistance gene, the broad detoxification spectrum of *Fhb7* through de-epoxidation of trichothecenes suggests the potential utility of the GST enzyme in the biomedicine, feed, and food industries in addition to reducing DON content in wheat grain.

HGT, the transfer of genes between non-mating species, is thought to occur frequently in prokaryotes, but much less so in eukaryotes (30). There is accumulating evidence illustrating instances of HGT events involving bacteria or the organellar genomes of another plant as donor (31). For instance, two *Agrobacterium* genes were found to be inserted in the genome (with transfer DNA borders) of a cultivated sweet potato [*Ipomoea batatas* (L.) Lam.], revealing a naturally occurring transgenic food crop (32). However, there is little evidence for HGT events involving nuclear DNA transmission from fungi or other eukaryotes, and such transmission has been thought to be insignificant (33). Fundamentally, our results highlight the roles that FP-HGT has had in shaping plant genomes, which advances the

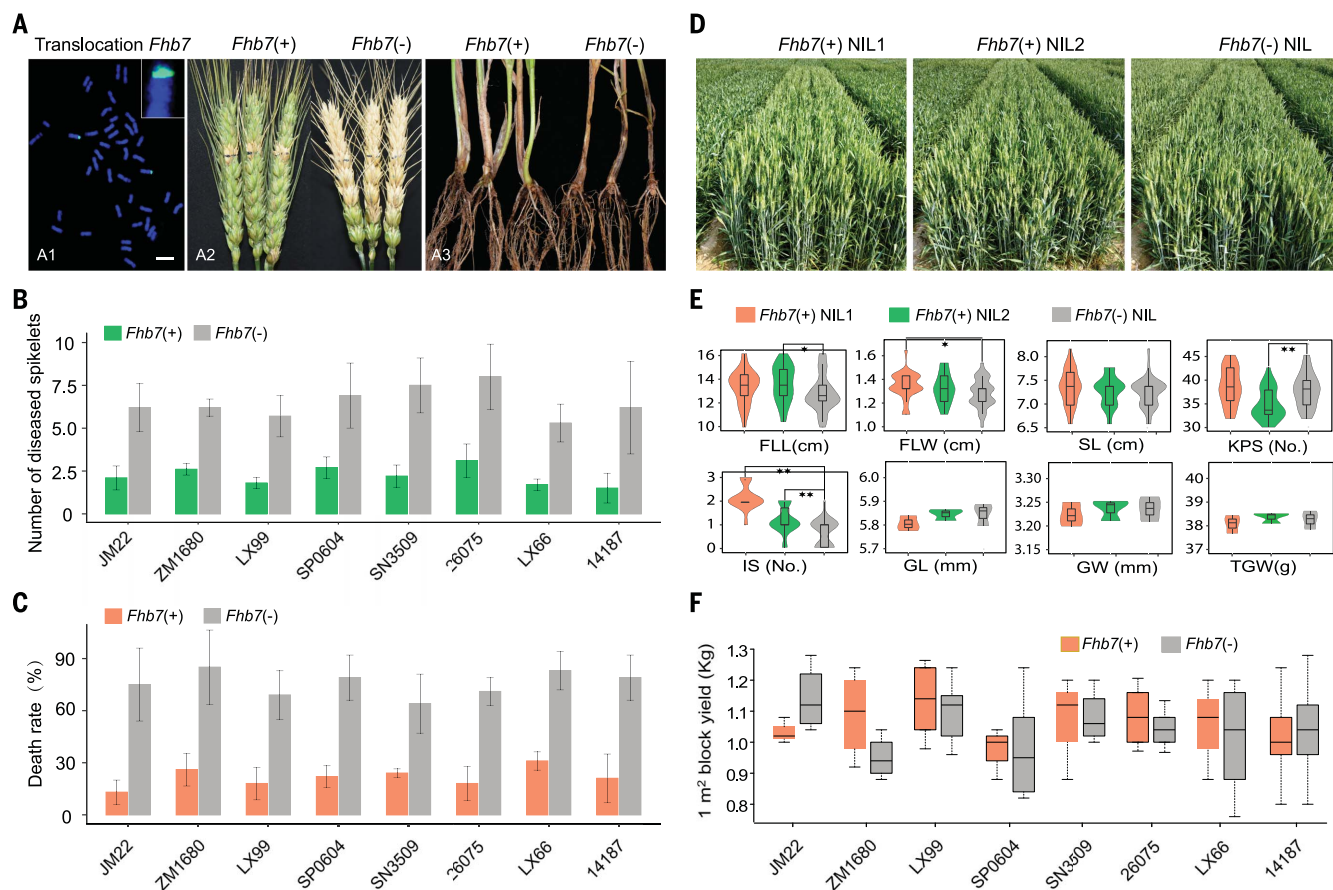


Fig. 3. Application prospects for *Fhb7* in wheat resistance breeding. (A) Genomic in situ hybridization analysis (left panel) showing a translocation of the distal region of 7E (containing *Fhb7*) from an E genome donor into wheat. Scale bar, 20 μ m. Also shown are images of *Fusarium*-infected spikes (middle panel) and crown rot (right panel) of LX99 NILs contrasting in *Fhb7*. (B) FHB resistance of *Fhb7* in eight different wheat genetic backgrounds evaluated at 21 d after inoculation in field conditions. (C) Crown rot phenotypes were recorded as the death ratio after growth in soil containing *F. pseudograminearum* at 30 days postinfection.

(D) Field plant photographs of two *Fhb7*(+) NILs and one *Fhb7*(-) NIL in the LX99 background. (E) Comparison of the yield traits among the two *Fhb7*(+) NILs and one *Fhb7*(-) NIL in the LX99 background evaluated in the 2017 field experiment. FLL, flag leaf length (cm); FLW, flag leaf width (cm); SL, spike length (cm); KPS, kernels per spike; IS, infertile spikelets; GL, grain length (mm); GW, grain width (mm); TGW, thousand grain weight (g). (F) Comparison of the grain yield among eight *Fhb7* translocation lines in different wheat genetic backgrounds. The grain yield was measured from a 1-m² plot in the 2017 and 2018 field experiments.

knowledge on disease resistance gene evolution and opens a new avenue for the identification of plant resistance genes.

The endophytic *Neotyphodium* and *Epichloë* fungi often form mutualistic symbiotic associations with forage grasses and offer hosts bioprotective benefits against pathogens and abiotic stresses, presumably owing to the fungus-mediated anabolism and catabolism of various natural product compounds (34). Here, we showed that the GST encoded by *Fhb7* is conserved in *Epichloë* species and is able to detoxify the trichothecene mycotoxins secreted by *Fusarium* species. Thus, transfer of this fungal gene into a plant genome could be beneficial to plants, perhaps even eliminating the need for the symbiotic association per se. The finding of *Fhb7*-mediated resistance to both FHB and crown rot diseases further emphasizes the importance of this HGT in benefiting the perennial *Th. elongatum*, which is

perhaps reflected by constitutive expression of *Fhb7* in all examined tissues. However, the molecular machinery that enabled the FP-HGT of *Fhb7* and the nature of the promoter evolution underlying the expression of *Fhb7* remain to be elucidated.

Methods summary

The *Th. elongatum* genome was first sequenced by Illumina short-read sequencing and was de novo assembled using the software package DeNovoMAGICTM3.0. PacBio SMRT long reads were used to fill the gaps in the assembly and Bionano optical maps were then used to correct and extend the scaffold sequences. The assembly was anchored into seven pseudochromosomes using Hi-C data. The assembly was validated using independent BAC sequences, genetic maps of related species, and commonly used software programs. Genes, repetitive DNA, and other genomic features

in the assembly were annotated to reveal the landscape of the species and to examine their relationship with wheat and other related species by in-depth comparative analyses. Genetic markers in the *Fhb7* region were developed by means of the reference genome sequence and used to screen recombinants for fine mapping to identify the *Fhb7* candidate gene. The candidate gene was functionally validated by virus-induced gene silencing, EMS-induced mutation, and transgenic approaches. FHB resistance was evaluated by inoculation of *Fusarium* conidial suspensions on wheat spikes, leaves, or crowns. LC-HRMS/MS analysis was used to infer the biochemical structure of trichothecene-glutathione adducts catalyzed by *Fhb7*. *Fhb7* was introgressed into diverse wheat backgrounds using distant hybridization and conventional breeding, and the presence of alien chromatin in wheat was validated by genomic in situ hybridization.

REFERENCES AND NOTES

- J. K. Haile *et al.*, *Fusarium* head blight in durum wheat: Recent status, breeding directions, and future research prospects. *Phytopathology* **109**, 1664–1675 (2019). doi: [10.1094/PHYTO-03-19-0095-RVW](https://doi.org/10.1094/PHYTO-03-19-0095-RVW); pmid: [31369363](https://pubmed.ncbi.nlm.nih.gov/31369363/)
- F. Trail, For blighted waves of grain: *Fusarium graminearum* in the postgenomics era. *Plant Physiol.* **149**, 103–110 (2009). doi: [10.1104/pp.108.129684](https://doi.org/10.1104/pp.108.129684); pmid: [19126701](https://pubmed.ncbi.nlm.nih.gov/19126701/)
- A. Hathout, S. Aly, Biological detoxification of mycotoxins: A review. *Ann. Microbiol.* **64**, 905–919 (2014). doi: [10.1007/s13213-014-0899-7](https://doi.org/10.1007/s13213-014-0899-7)
- Y. Hao, A. Rasheed, Z. Zhu, B. B. H. Wulff, Z. He, Harnessing wheat *Fhb1* for *Fusarium* resistance. *Trends Plant Sci.* **25**, 1–3 (2020). doi: [10.1016/j.tplants.2019.10.006](https://doi.org/10.1016/j.tplants.2019.10.006); pmid: [31679993](https://pubmed.ncbi.nlm.nih.gov/31679993/)
- E. S. Lagudah, S. G. Krattinger, A new player contributing to durable *Fusarium* resistance. *Nat. Genet.* **51**, 1070–1071 (2019). doi: [10.1038/s41588-019-0454-3](https://doi.org/10.1038/s41588-019-0454-3); pmid: [31253973](https://pubmed.ncbi.nlm.nih.gov/31253973/)
- G. Li *et al.*, Mutation of a histidine-rich calcium-binding-protein gene in wheat confers resistance to *Fusarium* head blight. *Nat. Genet.* **51**, 1106–1112 (2019). doi: [10.1038/s41588-019-0426-7](https://doi.org/10.1038/s41588-019-0426-7); pmid: [31182810](https://pubmed.ncbi.nlm.nih.gov/31182810/)
- Z. Su *et al.*, A deletion mutation in *TaHRC* confers *Fhb1* resistance to *Fusarium* head blight in wheat. *Nat. Genet.* **51**, 1099–1105 (2019). doi: [10.1038/s41588-019-0425-8](https://doi.org/10.1038/s41588-019-0425-8); pmid: [31182809](https://pubmed.ncbi.nlm.nih.gov/31182809/)
- N. Rawat *et al.*, Wheat *Fhb1* encodes a chimeric lectin with agglutinin domains and a pore-forming toxin-like domain conferring resistance to *Fusarium* head blight. *Nat. Genet.* **48**, 1576–1580 (2016). doi: [10.1038/ng.3706](https://doi.org/10.1038/ng.3706); pmid: [27776114](https://pubmed.ncbi.nlm.nih.gov/27776114/)
- L. Cui *et al.*, Development of perennial wheat through hybridization between wheat and wheatgrasses: A review. *Engineering* **4**, 507–513 (2018). doi: [10.1016/j.eng.2018.07.003](https://doi.org/10.1016/j.eng.2018.07.003)
- X. Zhang *et al.*, A genetic map of *Lophopyrum ponticum* chromosome 7E, harboring resistance genes to *Fusarium* head blight and leaf rust. *Theor. Appl. Genet.* **122**, 263–270 (2011). doi: [10.1007/s00122-010-1441-3](https://doi.org/10.1007/s00122-010-1441-3); pmid: [20830464](https://pubmed.ncbi.nlm.nih.gov/20830464/)
- L. L. Qi, M. O. Pumphrey, B. Friebe, P. D. Chen, B. S. Gill, Molecular cytogenetic characterization of alien introgressions with gene *Fhb3* for resistance to *Fusarium* head blight disease of wheat. *Theor. Appl. Genet.* **117**, 1155–1166 (2008). doi: [10.1007/s00122-008-0853-9](https://doi.org/10.1007/s00122-008-0853-9); pmid: [18712343](https://pubmed.ncbi.nlm.nih.gov/18712343/)
- C. Ceoloni *et al.*, Cytogenetic mapping of a major locus for resistance to *Fusarium* head blight and crown rot of wheat on *Thinopyrum elongatum* 7EL and its pyramiding with valuable genes from a *Th. ponticum* homoeologous arm onto bread wheat 7DL. *Theor. Appl. Genet.* **130**, 2005–2024 (2017). doi: [10.1007/s00122-017-2939-8](https://doi.org/10.1007/s00122-017-2939-8); pmid: [28656363](https://pubmed.ncbi.nlm.nih.gov/28656363/)
- J. Guo *et al.*, High-density mapping of the major FHB resistance gene *Fhb7* derived from *Thinopyrum ponticum* and its pyramiding with *Fhb1* by marker-assisted selection. *Theor. Appl. Genet.* **128**, 2301–2316 (2015). doi: [10.1007/s00122-015-2586-x](https://doi.org/10.1007/s00122-015-2586-x); pmid: [26220223](https://pubmed.ncbi.nlm.nih.gov/26220223/)
- H. Li, X. Wang, *Thinopyrum ponticum* and *Th. intermedium*: The promising source of resistance to fungal and viral diseases of wheat. *J. Genet. Genomics* **36**, 557–565 (2009). doi: [10.1016/S1673-8527\(08\)60147-2](https://doi.org/10.1016/S1673-8527(08)60147-2); pmid: [19782957](https://pubmed.ncbi.nlm.nih.gov/19782957/)
- X. Shen, H. Ohm, Molecular mapping of *Thinopyrum*-derived *Fusarium* head blight resistance in common wheat. *Mol. Breed.* **20**, 131–140 (2007). doi: [10.1007/s11032-007-9079-9](https://doi.org/10.1007/s11032-007-9079-9)
- R. P. Singh, R. A. McIntosh, Genetics of resistance to *Puccinia graminis* tritici in 'Chris' and 'W3746' wheats. *Theor. Appl. Genet.* **73**, 846–855 (1987). doi: [10.1007/BF00289389](https://doi.org/10.1007/BF00289389); pmid: [24241294](https://pubmed.ncbi.nlm.nih.gov/24241294/)
- R. M. Waterhouse *et al.*, BUSCO Applications from quality assessments to gene prediction and phylogenomics. *Mol. Biol. Evol.* **35**, 543–548 (2018). doi: [10.1093/molbev/msx319](https://doi.org/10.1093/molbev/msx319); pmid: [29220515](https://pubmed.ncbi.nlm.nih.gov/29220515/)
- S. Ou, J. Chen, N. Jiang, Assessing genome assembly quality using the LTR Assembly Index (LAI). *Nucleic Acids Res.* **46**, e126 (2018). doi: [10.1093/nar/gky730](https://doi.org/10.1093/nar/gky730); pmid: [30107434](https://pubmed.ncbi.nlm.nih.gov/30107434/)
- T. Kantarski *et al.*, Development of the first consensus genetic map of intermediate wheatgrass (*Thinopyrum intermedium*) using genotyping-by-sequencing. *Theor. Appl. Genet.* **130**, 137–150 (2017). doi: [10.1007/s00122-016-2799-7](https://doi.org/10.1007/s00122-016-2799-7); pmid: [27738715](https://pubmed.ncbi.nlm.nih.gov/27738715/)
- B. S. Gaut, B. R. Morton, B. C. McCaig, M. T. Clegg, Substitution rate comparisons between grasses and palms: Synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcl*. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 10274–10279 (1996). doi: [10.1073/pnas.93.19.10274](https://doi.org/10.1073/pnas.93.19.10274); pmid: [8816790](https://pubmed.ncbi.nlm.nih.gov/8816790/)
- J. Guo *et al.*, Molecular and cytological comparisons of chromosomes 7e₁, 7e₂, 7E^a, and 7E^b derived from *Thinopyrum*. *Cytogenet. Genome Res.* **145**, 68–74 (2015). doi: [10.1159/000381838](https://doi.org/10.1159/000381838); pmid: [25968454](https://pubmed.ncbi.nlm.nih.gov/25968454/)
- X. Shen, L. Kong, H. Ohm, *Fusarium* head blight resistance in hexaploid wheat (*Triticum aestivum*)-*Lophopyrum* genetic lines and tagging of the alien chromatin by PCR markers. *Theor. Appl. Genet.* **108**, 808–813 (2004). doi: [10.1007/s00122-003-1492-9](https://doi.org/10.1007/s00122-003-1492-9); pmid: [14628111](https://pubmed.ncbi.nlm.nih.gov/14628111/)
- S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. J. Lipman, Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990). doi: [10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2); pmid: [2231712](https://pubmed.ncbi.nlm.nih.gov/2231712/)
- M. Morel, A. A. Ngadin, M. Droux, J. P. Jacquot, E. Gelhaye, The fungal glutathione S-transferase system. Evidence of new classes in the wood-degrading basidiomycete *Phanerochaete chrysosporium*. *Cell. Mol. Life Sci.* **66**, 3711–3725 (2009). doi: [10.1007/s00018-009-0104-5](https://doi.org/10.1007/s00018-009-0104-5); pmid: [19662500](https://pubmed.ncbi.nlm.nih.gov/19662500/)
- S. Walter, P. Nicholson, F. M. Doohan, Action and reaction of host and pathogen during *Fusarium* head blight disease. *New Phytol.* **185**, 54–66 (2010). doi: [10.1111/j.1469-8137.2009.03041.x](https://doi.org/10.1111/j.1469-8137.2009.03041.x); pmid: [19807873](https://pubmed.ncbi.nlm.nih.gov/19807873/)
- B. Kluger *et al.*, Stable isotopic labelling-assisted untargeted metabolic profiling reveals novel conjugates of the mycotoxin deoxynivalenol in wheat. *Anal. Bioanal. Chem.* **405**, 5031–5036 (2013). doi: [10.1007/s00216-012-6483-8](https://doi.org/10.1007/s00216-012-6483-8); pmid: [23086087](https://pubmed.ncbi.nlm.nih.gov/23086087/)
- S. A. Gardiner *et al.*, Transcriptome analysis of the barley-deoxynivalenol interaction: Evidence for a role of glutathione in deoxynivalenol detoxification. *Mol. Plant Microbe Interact.* **23**, 962–976 (2010). doi: [10.1094/MPMI-23-7-0962](https://doi.org/10.1094/MPMI-23-7-0962); pmid: [20521958](https://pubmed.ncbi.nlm.nih.gov/20521958/)
- A. Stanic *et al.*, Characterization of deoxynivalenol–glutathione conjugates using nuclear magnetic resonance spectroscopy and liquid chromatography–high-resolution mass spectrometry. *J. Agric. Food Chem.* **64**, 6903–6910 (2016). doi: [10.1021/acs.jafc.6b02853](https://doi.org/10.1021/acs.jafc.6b02853); pmid: [27548277](https://pubmed.ncbi.nlm.nih.gov/27548277/)
- F. Berthiller *et al.*, Occurrence of deoxynivalenol and its 3- β -D-glucoside in wheat and maize. *Food Addit. Contam. Part A Chem. Anal. Control Expo. Risk Assess.* **26**, 507–511 (2009). doi: [10.1080/02652030802555668](https://doi.org/10.1080/02652030802555668); pmid: [19680925](https://pubmed.ncbi.nlm.nih.gov/19680925/)
- F. Husnik, J. P. McCutcheon, Functional horizontal gene transfer from bacteria to eukaryotes. *Nat. Rev. Microbiol.* **16**, 67–79 (2018). doi: [10.1038/nrmicro.2017.137](https://doi.org/10.1038/nrmicro.2017.137); pmid: [29176581](https://pubmed.ncbi.nlm.nih.gov/29176581/)
- P. J. Keeling, J. D. Palmer, Horizontal gene transfer in eukaryotic evolution. *Nat. Rev. Genet.* **9**, 605–618 (2008). doi: [10.1038/nrg2386](https://doi.org/10.1038/nrg2386); pmid: [18591983](https://pubmed.ncbi.nlm.nih.gov/18591983/)
- T. Kyndt *et al.*, The genome of cultivated sweet potato contains *Agrobacterium* T-DNAs with expressed genes: An example of a naturally transgenic food crop. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 5844–5849 (2015). doi: [10.1073/pnas.1419685112](https://doi.org/10.1073/pnas.1419685112); pmid: [25902487](https://pubmed.ncbi.nlm.nih.gov/25902487/)
- H. Shinozuka *et al.*, Horizontal transfer of a β -1,6-glucanase gene from an ancestral species of fungal endophyte to a cool-season grass host. *Sci. Rep.* **7**, 9024 (2017). doi: [10.1038/s41598-017-07886-2](https://doi.org/10.1038/s41598-017-07886-2); pmid: [28831055](https://pubmed.ncbi.nlm.nih.gov/28831055/)
- A. Tanaka, D. Takemoto, T. Chujo, B. Scott, Fungal endophytes of grasses. *Curr. Opin. Plant Biol.* **15**, 462–468 (2012). doi: [10.1016/j.pbi.2012.03.007](https://doi.org/10.1016/j.pbi.2012.03.007); pmid: [22465162](https://pubmed.ncbi.nlm.nih.gov/22465162/)
- L. Zhao *et al.*, Cloning and characterization of a specific UDP-glycosyltransferase gene induced by DON and *Fusarium graminearum*. *Plant Cell Rep.* **37**, 641–652 (2018). doi: [10.1007/s00299-018-2257-x](https://doi.org/10.1007/s00299-018-2257-x); pmid: [29372381](https://pubmed.ncbi.nlm.nih.gov/29372381/)

ACKNOWLEDGMENTS

We thank Q. Song (Lanzhou University) and X. Zhang (University of Louisville) for advice on entophytic fungi and biochemical analysis of trichothecenes, X. L. Zhang (Northeast Forestry University) for providing some marker sequences of the Triticeae E genome, and Y. Liang and J. Yu (Shandong Agricultural University) for providing the *Fusarium* stains. **Funding:** This work was supported by the National Natural Science Foundation of China (31520103911, 31871610, and 31901492), the National Key Research and Development Program (2016YFD0100102-2), the National Key Program on Transgenic Research from the Ministry of Agriculture of China (2016ZX08002003-002 and 2016ZX08009-003), and the Agricultural Variety Improvement Project of Shandong Province (2019LZGC016). **Author contributions:** H.W. and L.K. designed the project. S.S.X., G.B., E.N., C.G., and H.O. supervised the project. S.S., K.W., L.C., X.F., and F.N. performed bioinformatics analysis. W.G., L.Z., B.H., Z.L., S.S.X., J.G., M.L., P.S., X.F.L., G.W., C.B., W.Z., X.C., J.W., L.D., W.C., W.L., G.X., J.Z., Y.H., Y.X., Y.G., W.J.L., Y.L., H.Y., J.L., X.L., Y.Z., and X.W. conducted experiments. X.M. and A.L. performed the field work. H.W., S.S., G.B., and L.K. wrote the paper with input from all authors. **Competing interests:** The authors declare no competing interests. Patents with application nos. 2020101464009 and 202010146399X are pending. **Data and materials availability:** All data are available in the manuscript, the supplementary materials, or at publicly accessible repositories. These data in the public repositories include all raw reads and assembled sequence data for *Th. elongatum* in NCBI under BioProjectID PRJNA540081, the assembly and annotation data of *Th. elongatum*, and the draft genomes of 7E/7D substitution lines in the Genome Warehouse in BIG Data Center under accession numbers GWHABKY000000000, GWHABLF000000000, and GWHABLE000000000, which are accessible at <https://bigd.big.ac.cn/gwh>. All materials are available from L. Kong upon request.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6493/eaba5435/suppl/DC1
Materials and Methods
Figs. S1 to S40
Tables S1 to S23
Captions for Data S1 to S3
References (36–93)
MDAR Reproducibility Checklist
Data S1 to S3

[View/request a protocol for this paper from Bio-protocol.](#)

12 December 2019; accepted 26 March 2020
Published online 9 April 2020
10.1126/science.aba5435

RESEARCH ARTICLE

WATER RESOURCES

Global threat of arsenic in groundwater

Joel Podgorski^{1,2*} and Michael Berg^{1,3*}

Naturally occurring arsenic in groundwater affects millions of people worldwide. We created a global prediction map of groundwater arsenic exceeding 10 micrograms per liter using a random forest machine-learning model based on 11 geospatial environmental parameters and more than 50,000 aggregated data points of measured groundwater arsenic concentration. Our global prediction map includes known arsenic-affected areas and previously undocumented areas of concern. By combining the global arsenic prediction model with household groundwater-usage statistics, we estimate that 94 million to 220 million people are potentially exposed to high arsenic concentrations in groundwater, the vast majority (94%) being in Asia. Because groundwater is increasingly used to support growing populations and buffer against water scarcity due to changing climate, this work is important to raise awareness, identify areas for safe wells, and help prioritize testing.

The natural, or geogenic, occurrence of arsenic in groundwater is a global problem with wide-ranging health effects for humans and wildlife. Because it is toxic and does not serve any beneficial metabolic function, inorganic arsenic (the species present in groundwater) can lead to disorders of the skin and vascular and nervous systems,

as well as cancer (1, 2). The major source of inorganic arsenic in the diet is through arsenic-contaminated water; although ingestion through food, particularly rice, represents another important route of exposure (3). As a consequence, the World Health Organization (WHO) has set a guideline arsenic concentration of 10 $\mu\text{g}/\text{liter}$ in drinking water (4).

At least trace amounts of arsenic occur in virtually all rocks and sediments around the world (5). However, in most of the large-scale cases of geogenic arsenic contamination in groundwater, arsenic accumulates in aquifers composed of recently deposited alluvial sediments. Under anoxic conditions, arsenic is released from the microbial and/or chemical reductive dissolution of arsenic-bearing iron(III) minerals in the aquifer sediments (6–9). Un-

der oxidizing, high-pH conditions, arsenic can also desorb from iron and aluminum hydroxides (10). Furthermore, aquifers in flat-lying sedimentary sequences generally have a small hydraulic gradient, causing groundwater to flow slowly. This longer groundwater residence time allows dissolved arsenic to accumulate and its concentration to increase. Other processes responsible for arsenic release into groundwater include oxidation of arsenic-bearing sulfide minerals as well as release from arsenic-enriched geothermal deposits.

That arsenic is generally not included in the standard suite of tested water quality parameters (11) and is not detected by the human senses means that arsenic is regularly being discovered in new areas. Since one of the greatest occurrences of geogenic groundwater arsenic was discovered in 1993 in the Bengal delta (5, 12, 13), high arsenic concentrations have been detected all around the world, with hot spots including Argentina (14–17), Cambodia (18, 19), China (20–22), India (23–25), Mexico (26, 27), Pakistan (28, 29), the United States (30, 31), and Vietnam (32, 33).

To help identify areas likely to contain high concentrations of arsenic in groundwater, several researchers have used statistical learning methods to create arsenic prediction maps based on available datasets of measured arsenic concentrations and relevant geospatial parameters. Previous studies have focused on Burkina Faso (34), China (21, 35), South Asia (29, 36), Southeast Asia (37), the United States (31, 38, 39), and the Red River delta in Vietnam (33), as well as sedimentary basins around the world (40). The predictor variables used in these studies generally include various climate and soil parameters, geology, and topography (table S3).

¹Department of Water Resources and Drinking Water, Eawag, Swiss Federal Institute of Aquatic Science and Technology, 8600 Dübendorf, Switzerland. ²Department of Earth and Environmental Sciences, University of Manchester, Manchester M13 9PL, UK. ³UNESCO Chair on Groundwater Arsenic within the 2030 Agenda for Sustainable Development and School of Civil Engineering and Surveying, University of Southern Queensland, Toowoomba, QLD 4350, Australia.
*Corresponding author. Email: joel.podgorski@eawag.ch (J.P.); michael.berg@eawag.ch (M.B.)

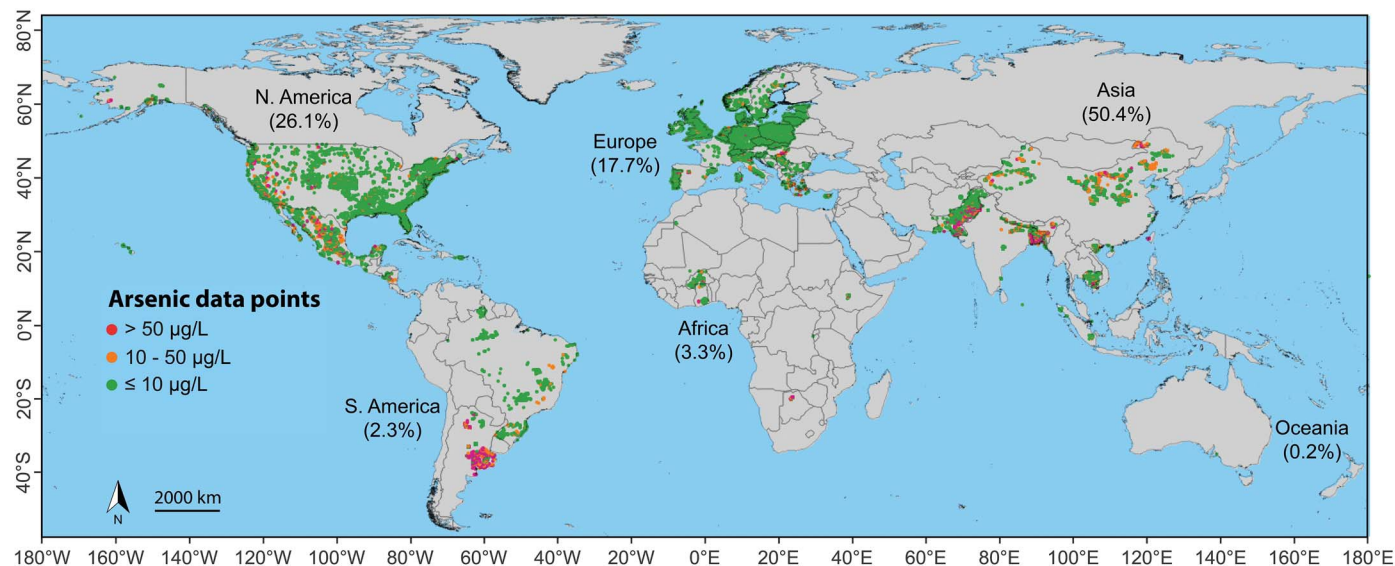


Fig. 1. Arsenic concentrations, excluding those known to originate from a depth greater than 100 m. Values are from the sources listed in table S1. The geographical distribution of data is indicated by continent.

Taking advantage of the increasing availability of high-resolution datasets of relevant environmental parameters, we use statistical learning to model what to our knowledge is the most spatially extensive compilation of arsenic measurements in groundwater assembled, which makes a global model possible. To focus on health risks, we consider the probability of arsenic in groundwater exceeding the WHO guideline. For this, we have chosen the random forest method, which our preliminary tests showed to be highly effective in addressing this classification problem. We use the resulting model to produce the most accurate and detailed global prediction map to date of geogenic groundwater arsenic, which can be used to help identify previously unknown areas of arsenic contamination as well as more clearly

delineate the scope of this global problem and considerably increase awareness.

Results

Random forest modeling

We aggregated data from nearly 80 studies of arsenic in groundwater (see table S1 for references and statistics) into a single dataset ($n > 200,000$). Averaging into 1-km² pixels resulted in more than 55,000 arsenic data points for use in modeling based on groundwater samples not known to originate from greater than 100-m depth (Fig. 1).

To create the simplest and most accurate model, an initial set of 52 potentially relevant environmental predictor variables was iteratively reduced in consideration of their relative importance and impact on the accuracy

of a succession of random forest models. The final selection of 11 predictor variables (table S2) includes several soil parameters (topsoil clay, subsoil sand, pH, and fluvisols), all of the climate variables (precipitation, actual and potential evapotranspiration, and combinations thereof, as well as temperature), and the topographic wetness index. By contrast, none of the geology variables proved to be statistically important. This is not to imply that geology does not play a role in geogenic arsenic accumulation, but rather that the particular geology variables tested were not as relevant as the other variables. This may be due to the coarse nature of the geological maps, which are standardized for the entire world. Although the number of predictor variables was reduced by nearly 80%, both the area

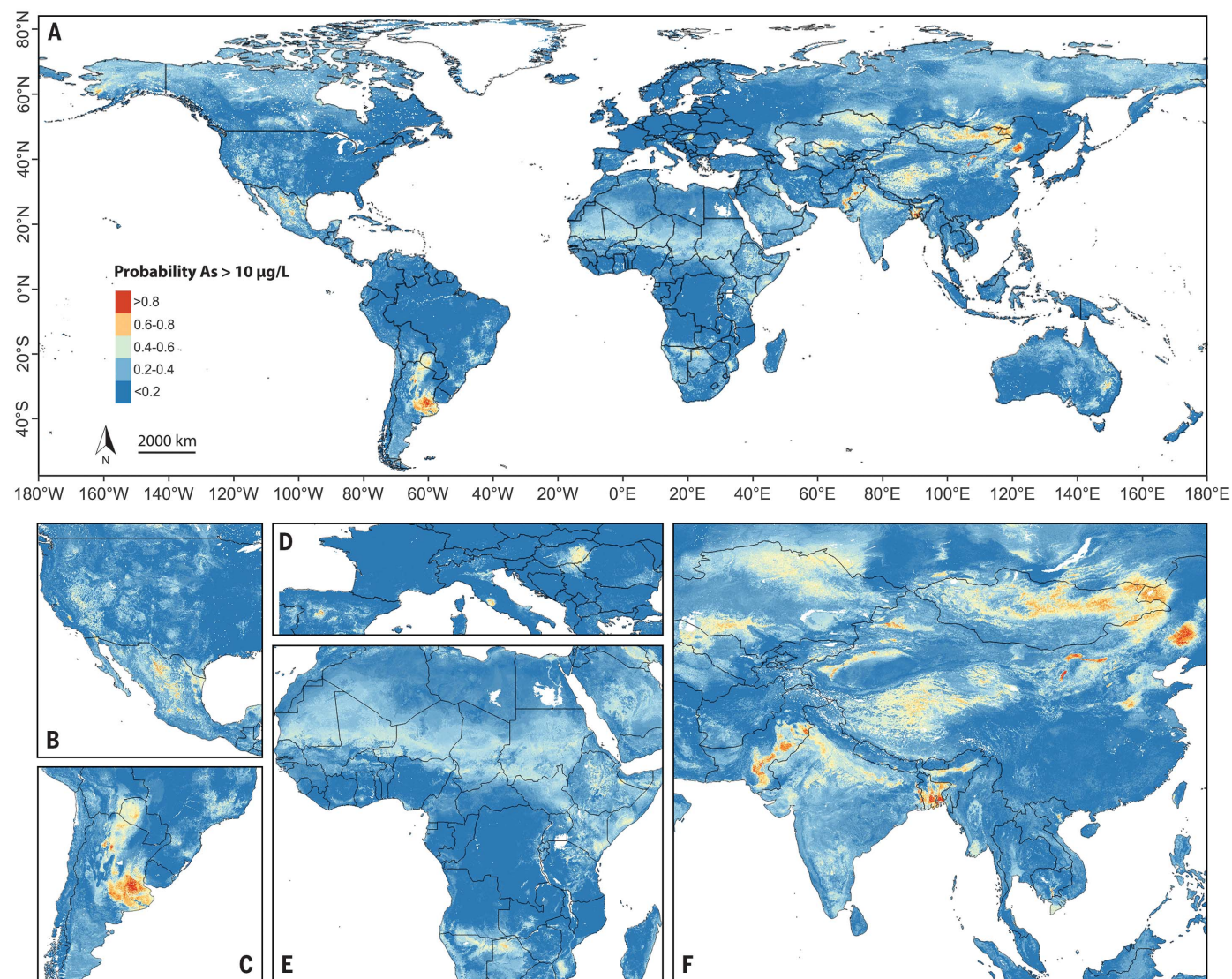


Fig. 2. Global prediction of groundwater arsenic. (A to F) Modeled probability of arsenic concentration in groundwater exceeding 10 µg/liter for the entire globe (A) along with zoomed-in sections of the main more densely populated affected areas (B) to (F). The model is based on the arsenic data points in Fig. 1 and the predictor variables in table S2. Figs. S2 to S8 provide more detailed views of the prediction map.

under the curve (AUC, 0.89) and Cohen’s kappa statistic (0.55) remained unchanged.

The final random forest model was created based on the compiled global dataset of high and low arsenic concentrations along with the 11 predictor variables. The standard number of variables to be made available at each branch of each tree is between three and four (see methods). Because our tests showed the value of three performing better than four and higher values (though error and performance rates varied only within ~1%), we set this parameter to three. The global map produced from this model is displayed in Fig. 2A along with more detailed views of the more populated affected continental regions shown in Fig. 2, B to F. It indicates the probability of the concentration of arsenic in groundwater in a given 1-km² cell exceeding 10 µg/liter. The uncertainty of the model is inherent in the probabilities themselves, because they are simply the average of the votes or predictions of high or low values of each of the 10,001 trees grown. That is, each tree casts a vote of 0 or 1 (“no” or “yes” to As > 10 µg/liter) for each cell based on the values of the predictor variables in that cell. Figures S2

to S8 also provide more detailed views of the prediction map for each of the inhabited continents.

The importance of each of the 11 predictor variables in terms of mean decrease in accuracy and mean decrease in the Gini index is listed in fig. S1. Relative to the initial set of 52 variables, the values of these two statistics for most of the 11 final predictor variables appear to fall within a fairly narrow range, indicating comparable importance. Exceptions include fluvisols and soil pH, which have somewhat greater importance, and temperature, which, according to both statistics, is the least important of the 11 variables. Soil pH was also found to be an important predictor variable in arid, oxidizing environments in Pakistan (29). Although widespread arsenic dissolution occurs in Holocene fluvial sediments (5–7, 9, 37), this geological epoch has not been consistently mapped around the world. However, the global dataset of fluvisols provides a very suitable alternative (29), which may even be more appropriate because fluvisols by definition encompass recent fluvial sediments and not, for example, aeolian Holocene

sediments that are generally not relevant for arsenic release. The generally high model importance of climate variables, as evidenced by them all being selected for the final model, highlights the strong control that climate has on arsenic release in aquifers. In particular, precipitation and evapotranspiration have a direct role in creating conditions conducive for arsenic release under reducing conditions (e.g., waterlogged soils) as well as high aridity associated with oxidizing, high-pH conditions.

The performance of the random forest model on the test dataset (20% of the data, which was randomly selected while maintaining the relative distribution of high and low values) is summarized in the confusion matrix in Table 1. Despite a prevalence of high values (>10 µg/liter) of only 22% in the dataset, the model performs well in predicting both high values (sensitivity: 0.79) and low values (specificity: 0.85) at a probability cutoff of 0.50. The average of these two figures, known as balanced accuracy, is correspondingly high at 0.82. Likewise, the model’s AUC, which considers the full range of possible cutoffs, has a very high value of 0.89 with the test dataset (Table 1). For comparison, the AUC of a random forest using all 52 original predictor variables is also 0.89.

The model was also tested on a dataset of more than 49,000 arsenic data points originating from known depths greater than 100 m (average 562 m, standard deviation 623 m). Although the model was not trained on any measurements from these depths and the fact that only surface parameters were used as predictor variables, the model nevertheless performed quite well in predicting the arsenic concentrations of these deep groundwater sources, as evidenced by an AUC of 0.77.

Regions and populations at risk

Areas predicted to have high arsenic concentrations in groundwater exist on all continents, with most being located in Central, South, and Southeast Asia; parts of Africa; and North and South America (Fig. 2 and figs. S2 to S8). Known areas of groundwater arsenic contamination are generally well captured by the global arsenic prediction map, for example, parts of the western United States, central Mexico, Argentina, the Pannonian Basin, Inner Mongolia, the Indus Valley, the Ganges-Brahmaputra delta, and the Mekong River and Red River deltas. Areas of increased arsenic hazard where little concentration data exist include parts of Central Asia, particularly Kazakhstan, Mongolia, and Uzbekistan; the Sahel region; and broad areas of the Arctic and sub-Arctic. Of these, the Central Asian hazard areas are better constrained, as evidenced by higher probabilities.

Probability threshold values of 0.57 from the sensitivity-specificity comparison and 0.72 from the positive predictive value (PPV)–negative

Table 1. Confusion matrix and other statistics summarizing the results of applying the random forest model to the test dataset at a probability cutoff of 0.50.	
Model output	Value
Predicted As ≤ 10 µg/liter	
Measured As ≤ 10 µg/liter	7710
Measured As > 10 µg/liter	555
Predicted As > 10 µg/liter	
Measured As ≤ 10 µg/liter	1394
Measured As > 10 µg/liter	2037
Sensitivity	0.79
Specificity	0.85
PPV	0.59
NPV	0.93
Prevalence	0.22
Balanced accuracy	0.82
Cohen’s kappa	0.55
AUC	0.89

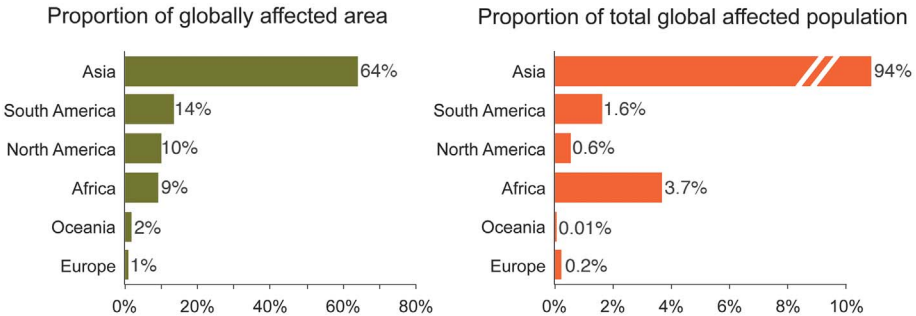


Fig. 3. Proportions of land area and population potentially affected by arsenic concentrations in groundwater exceeding 10 µg/liter by continent.

predictive value (NPV) comparison were found using the full dataset (combined training and test datasets) of arsenic concentrations. The proportions of high modeled arsenic hazard by continent associated with each of these probabilities are shown in Fig. 3. Global maps of the potentially affected population in the risk areas, as determined by these two thresholds, are shown in Fig. 4. As described in the methods, these maps were then used to estimate the population potentially affected by drinking groundwater with arsenic concentrations exceeding 10 $\mu\text{g}/\text{liter}$.

The resulting global arsenic risk assessment indicates that about 94 million to 220 million people around the world (of which 85 to 90% are in South Asia) are potentially exposed to high concentrations of arsenic in groundwater from their domestic water supply (tables S4 and S5). This range is consistent with the previous most comprehensive literature compilations, that is, 140 million people (41) and 225 million people (42). Household groundwater-use statistics were not available for ~6 to 8% of the affected countries (depending on the cutoff), for which the less detailed statistics derived from the AQUASTAT database of the Food and Agriculture Organization of the United Nations were used instead (see methods for details). To determine the amount of error that using these more general groundwater-use statistics might introduce to the overall population figures, the global potentially affected populations were recalculated with these countries' (those lacking household groundwater-use statistics) groundwater-use rates set to the extreme values of 0 and 100%. Because this applied to relatively few countries and arsenic-affected areas, doing so affected the overall global population figures by an inconsequential amount ($\pm 0.1\%$), indicating that using the AQUASTAT groundwater-use rates, where necessary, is an acceptable approximation.

This estimate of risk takes into account only the proportion of households utilizing unprocessed groundwater and assumes uniform rates throughout the urban and nonurban areas of each country. The uncertainties of these rates are unknown. The population in each cell was reduced by the uncertainty of the cell's prediction, which is justified based on the heterogeneity inherent in the accumulation of arsenic in an aquifer, which is generally at a much finer scale than that of the 1- km^2 resolution of the arsenic hazard map. Because the arsenic prediction for a cell represents the average outcome for that cell, we can take the modeled probability as a first-order approximation of the proportion of an aquifer in that cell containing high arsenic concentrations. Only cells exceeding the probability threshold (i.e., 0.57 or 0.72) were considered. The global estimate of 94 million to 220 million people potentially affected by consuming arsenic-contaminated groundwater is

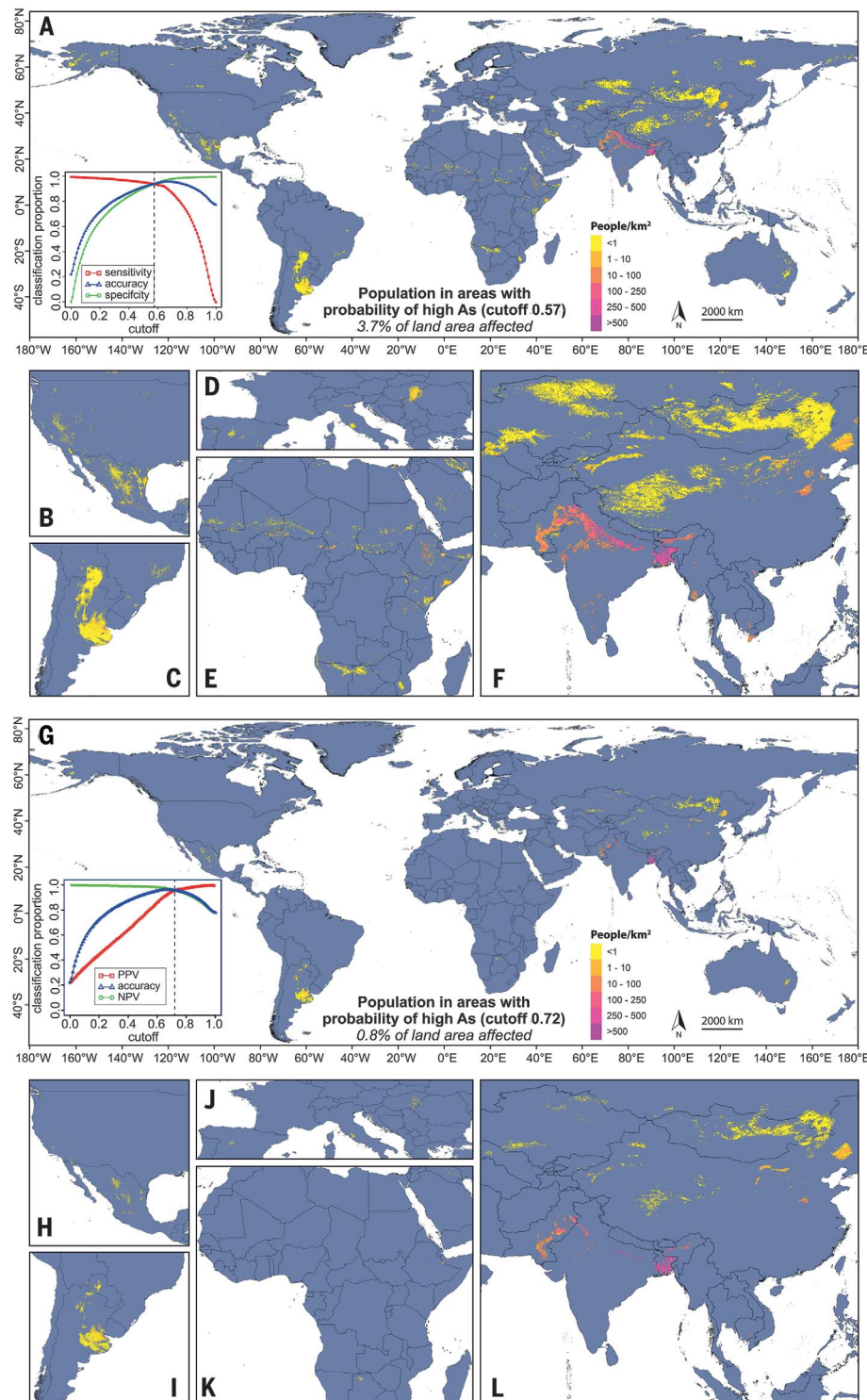


Fig. 4. Estimated population at risk. (A to L) Population in risk areas potentially containing aquifers with arsenic concentrations $>10 \mu\text{g}/\text{liter}$ using probability cutoffs of 0.57 (A), at which sensitivity and specificity are equal [inset in (A)] as applied to the full (training and test) dataset, and 0.72 (G), at which PPV and NPV are equal [inset in (G)] using the full dataset. The detailed areas of Fig. 2 are also repeated here for both models (B) to (F) and (H) to (L).

broken down by continent and country in tables S4 and S5, respectively, and represents the most accurate and consistent global estimate available.

Discussion

The accuracy of the global groundwater arsenic prediction model presented here, as indicated, for example, with an AUC of 0.89 calculated with the test dataset, exceeds that found in previous arsenic prediction studies (table S3). The dominance of climate and soil parameters in the final model is indicative of their direct influence or at least strong association with the processes of arsenic accumulation in groundwater.

With respect to previous arsenic prediction maps of global sedimentary basins (40, 43), the new model represents a substantial advancement on a few different levels. First, the new model presented here provides predictions for all areas of the inhabited continents, whereas the previous first-generation statistical model covered only about half of the land areas. In addition, a 10-fold increase in measurement points has allowed arsenic concentrations to be incorporated from many more areas of the globe. The greatly expanded availability and quality of global predictor datasets over the past 10 years has enabled new variables to be considered, such as soil type (e.g., fluvisols), as well as provided a 10- to 60-fold greater spatial resolution (i.e., 30 arc-sec versus 5 to 30 arc-min). However, the presence of high arsenic in groundwater at a given location is of course predicated on the existence of an aquifer in the first place, which may not be so in the case of unfractured solid rock, steep terrain, or very dry conditions. Models are only as good as the data on which they are based. As accurate as the new arsenic model is, it could be further improved as more arsenic data and more detailed predictor datasets come into existence.

Particularly in sedimentary aquifers, arsenic concentration is often highly dependent on depth, that is, on specific sedimentary sequences that differ in the concentration of arsenic in sediments as well as the geochemical conditions conducive to arsenic release. To better characterize this relationship in a given sedimentary basin, detailed depth information of groundwater samples would need to be incorporated in a separate basin-level study. Unfortunately, it is not feasible in a global-scale study to account for all of the diversity of the sedimentary basins of the world, especially because depth information of groundwater samples is often not available. As such, we have relied on a statistical analysis of model performance against depth ranges of samples (where present) to determine model sensitivity to depth.

Our approach in the risk assessment of potentially affected populations is relatively dis-

cerning and/or conservative. As such, the resulting population estimates may in some cases be lower than those found in earlier studies. One reason for this is that we used country-specific statistics of rural and urban domestic groundwater usage, which allowed us to subtract the proportion of the population that uses surface water, tap water, or other sources. This was not the case, for example, in a previous study of China that estimated that 19.6 million people were affected in the country (21), whereas our estimate is considerably lower at 4.3 million to 12.1 million. Furthermore, we consider only areas in which the probability of high arsenic exceeds the statistically determined cutoffs, that is, 0.57 and 0.72. Taking the United States as an example, applying this criterion left only 0.2 to 2% of the area of the country over which to sum the potentially affected population (≤ 0.21 million, this study). In a previous arsenic risk assessment of the United States (31), the entire country was used to estimate affected population (2.1 million), that is, not only the high-risk areas.

The actual proportion of groundwater usage varies spatially throughout a country, and so more detailed usage statistics beyond only urban versus rural would improve the accuracy of a risk assessment. In addition, more groundwater samples (ideally including depth information) from areas that currently have poor coverage would benefit future modeling efforts by allowing the model to be better adapted to those areas.

The presented arsenic probability maps should be used as a guide to further groundwater arsenic testing, for example, in Central Asia, the Sahel, and other regions of Africa. Only actual groundwater quality testing can definitively determine the suitability of groundwater with respect to arsenic, particularly because of small-scale (<1 km) aquifer heterogeneities that cannot be modeled with existing global datasets (9, 44). The hazard maps highlight areas at risk and provide a basis for targeted surveys, which continue to be important. The already large number of people potentially affected can be expected to increase as groundwater use expands with a growing population and increasing irrigation, especially in the light of water scarcity associated with warmer and drier conditions related to climate change. The maps can also help aid mitigation measures, such as awareness raising, coordination of government and financial support, health intervention programs, securing alternative drinking water resources, and arsenic removal options tailored to the local groundwater conditions as well as social setting.

REFERENCES AND NOTES

1. A. H. Smith, E. O. Lingas, M. Rahman, *Bull. World Health Organ.* **78**, 1093–1103 (2000).
2. M. F. Hughes, *Toxicol. Lett.* **133**, 1–16 (2002).
3. D. Mondal et al., *Environ. Geochem. Health* **32**, 463–477 (2010).

4. H. G. Gorchev, G. Ozolins, *WHO Chron.* **38**, 104–108 (1984).
5. P. Smedley, D. Kinniburgh, *Appl. Geochem.* **17**, 517–568 (2002).
6. R. Nickson et al., *Nature* **395**, 338–338 (1998).
7. J. McArthur, P. Ravenscroft, S. Safiulla, M. Thirlwall, *Water Resour. Res.* **37**, 109–117 (2001).
8. M. Berg et al., *Chem. Geol.* **249**, 91–112 (2008).
9. S. Fendorf, H. A. Michael, A. van Geen, *Science* **328**, 1123–1127 (2010).
10. M. I. Litter et al., *Sci. Total Environ.* **676**, 756–766 (2019).
11. Y. Zheng, S. V. Flanagan, *Environ. Health Perspect.* **125**, 085002 (2017).
12. P. Bhattacharya, D. Chatterjee, G. Jacks, *Int. J. Water Resour. Dev.* **13**, 79–92 (1997).
13. A. van Geen et al., *Water Resour. Res.* **39**, 1140 (2003).
14. H. B. Niccoli, J. M. Suriano, M. A. Gomez Peral, L. H. Ferpozzi, O. A. Baleani, *Environ. Geol. Water Sci.* **14**, 3–16 (1989).
15. P. Smedley, H. Niccoli, D. Macdonald, A. Barros, J. Tullio, *Appl. Geochem.* **17**, 259–284 (2002).
16. M. Blarasin, A. Cabrera, E. Matteoda, paper presented at the XXXIII IAH – 7th ALHSD Congress, Zacatecas, Mexico, 11 to 15 October 2004.
17. M. Auge, G. E. Viale, L. Sierra, in *VIII Congreso Argentino de Hidrogeología: Aguas subterráneas recurso estratégico* (Editorial de la Universidad Nacional de La Plata, 2013), vol. 2, pp. 58–63.
18. M. Berg et al., *Environ. Sci. Technol.* **35**, 2621–2626 (2001).
19. J. Buschmann, M. Berg, C. Stengel, M. L. Sampson, *Environ. Sci. Technol.* **41**, 2146–2152 (2007).
20. P. Smedley, M. Zhang, G. Zhang, Z. Luo, *Appl. Geochem.* **18**, 1453–1477 (2003).
21. L. Rodríguez-Lado et al., *Science* **341**, 866–868 (2013).
22. Y. Zhou et al., *Appl. Geochem.* **77**, 116–125 (2017).
23. D. Chatterjee, R. Roy, B. Basu, *Environ. Geol.* **49**, 188–206 (2005).
24. B. Nath, D. Stüben, S. B. Mallik, D. Chatterjee, L. Charlet, *Appl. Geochem.* **23**, 977–995 (2008).
25. B. A. Shah, *Curr. Sci.* **98**, 1359–1365 (2010).
26. B. Planer-Friedrich, “Hydrogeological and hydrochemical investigations in the Rioverde Basin, Mexico,” thesis, Institute of Geology, University of Mining and Technology Freiberg (2000).
27. M. T. Alarcón-Herrera et al., *J. Hazard. Mater.* **262**, 960–969 (2013).
28. R. Nickson, J. McArthur, B. Shrestha, T. Kyaw-Myint, D. Lowry, *Appl. Geochem.* **20**, 55–68 (2005).
29. J. E. Podgorski et al., *Sci. Adv.* **3**, e1700935 (2017).
30. J. D. Ayotte, M. G. Nielsen, G. R. Robinson Jr., R. B. Moore, *Water Resour. Invest. Rep.* **99**, 4162 (1999).
31. J. D. Ayotte, L. Medalie, S. L. Qi, L. C. Backer, B. T. Nolan, *Environ. Sci. Technol.* **51**, 12443–12454 (2017).
32. M. Berg et al., *Sci. Total Environ.* **372**, 413–425 (2007).
33. L. H. Winkel et al., *Proc. Natl. Acad. Sci. U.S.A.* **108**, 1246–1251 (2011).
34. A. Bretzler et al., *Sci. Total Environ.* **584–585**, 958–970 (2017).
35. Q. Zhang et al., *J. Hazard. Mater.* **262**, 1147–1153 (2013).
36. S. Bindal, C. K. Singh, *Water Res.* **159**, 65–76 (2019).
37. L. Winkel, M. Berg, M. Amini, S. J. Hug, C. A. Johnson, *Nat. Geosci.* **1**, 536–542 (2008).
38. Q. Yang, H. B. Jung, R. G. Marwinney, C. W. Culbertson, Y. Zheng, *Environ. Sci. Technol.* **46**, 2080–2087 (2012).
39. N. Yang, L. H. Winkel, K. H. Johannesson, *Environ. Sci. Technol.* **48**, 5660–5666 (2014).
40. M. Amini et al., *Environ. Sci. Technol.* **42**, 3669–3675 (2008).
41. P. Ravenscroft, H. Brammer, K. Richards, *Arsenic Pollution: A Global Synthesis* (Wiley, 2009), vol. 28.
42. S. Murcott, *Arsenic Contamination in the World* (IWA Publishing, 2012).
43. P. Ravenscroft, “Predicting the global extent of arsenic pollution of groundwater and its potential impact on human health,” unpublished report prepared for UNICEF, December 2007.
44. Y. Zheng, *Curr. Environ. Health Rep.* **4**, 373–382 (2017).
45. J. Podgorski, M. Berg, Podgorski_and_Berg_2020, ERIC/open (2020); <http://dx.doi.org/10.25678/0001ZT>.

ACKNOWLEDGMENTS

We thank our colleagues A. Bretzler and C. Zurbügg (Eawag), A. Steiner and S. Piers de Raveschoot (SDC), and D. A. Polya and R. Wu (University of Manchester) for their support, as well as the many providers of data, which were an essential component of this work. **Funding:** We thank the Swiss Agency for Development and Cooperation (project nos. 7F-09010.01.01 and 7F-09963.01.01) for long-term support and cofunding

of this study, as well as a University of Manchester EPSRC IAA Impact Support Fund Award to D. A. Polya. **Author contributions:** J.P.: Methodology, Modeling, Writing–Original draft preparation; M.B.: Supervision, Writing–Reviewing and Editing. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The modeling data, code, and raster output maps are available at ERIC/open (45). Arsenic

concentration data points and hazard and risk maps are also available for viewing on the GIS-based Groundwater Assessment Platform (GAP). www.gapmaps.org.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6493/845/suppl/DC1
Methods

Figs. S1 to S11
Tables S1 to S6
References (46–127)

10 November 2019; accepted 3 April 2020
10.1126/science.aba1510

DEVICE TECHNOLOGY

Aligned, high-density semiconducting carbon nanotube arrays for high-performance electronics

Lijun Liu^{1*}, Jie Han^{1*}, Lin Xu¹, Jianshuo Zhou¹, Chenyi Zhao¹, Sujuan Ding^{2,3}, Huiwen Shi¹, Mengmeng Xiao¹, Li Ding¹, Ze Ma¹, Chuanhong Jin^{2,3}, Zhiyong Zhang^{1,2,4†}, Lian-Mao Peng^{1,2,4‡}

Single-walled carbon nanotubes (CNTs) may enable the fabrication of integrated circuits smaller than 10 nanometers, but this would require scalable production of dense and electronically pure semiconducting nanotube arrays on wafers. We developed a multiple dispersion and sorting process that resulted in extremely high semiconducting purity and a dimension-limited self-alignment (DLSA) procedure for preparing well-aligned CNT arrays (within alignment of 9 degrees) with a tunable density of 100 to 200 CNTs per micrometer on a 10-centimeter silicon wafer. Top-gate field-effect transistors (FETs) fabricated on the CNT array show better performance than that of commercial silicon metal oxide–semiconductor FETs with similar gate length, in particular an on-state current of 1.3 milliamperes per micrometer and a recorded transconductance of 0.9 millisiemens per micrometer for a power supply of 1 volt, while maintaining a low room-temperature subthreshold swing of <90 millivolts per decade using an ionic-liquid gate. Batch-fabricated top-gate five-stage ring oscillators exhibited a highest maximum oscillating frequency of >8 gigahertz.

The development of modern integrated circuits (ICs) has required scaling of field-effect transistors (FETs) to provide increased density, performance, and energy efficiency (1). Ultrathin semiconducting channels with high carrier mobility minimize the short-channel effect in aggressively scaled FETs (such as sub-10 nm technology nodes) (2). Single-walled carbon nanotubes (CNTs) can be 10 times as energy-efficient as conventional complementary metal oxide–semiconductor (CMOS) FETs because electron transport is ballistic, and CNTs have excellent electrostatic properties (2–5). Furthermore, prototype transistors built on individual CNTs with gate lengths as short as 5 nm outperform Si CMOS transistors in terms of both intrinsic performance and power consumption (5).

However, CNT FETs with real performance exceeding that of Si CMOS FETs have not been realized at similar technology nodes because CNT materials available for research are still far from ideal for electronics. As a building block for high-performance digital electronics, the extremely scaled CNT FET (with a channel width of several tens of nanometers; Fig. 1A) should contain multiple semiconducting CNTs in the channel to provide sufficient driving ability (2–6). A high-density aligned semiconducting CNT array is required as the channel material for fabricating large-scale ICs.

The ideal material system is well established to be aligned CNT (A-CNT) arrays with a well-defined and consistent 5- to 10-nm pitch (100 to 200 CNTs/μm), a semiconducting purity estimated to be >99.9999%, and a narrow diameter distribution around 1.5 nm (Fig. 1B) (3).

Although A-CNTs have been produced through chemical vapor deposition (CVD) growth of nanotubes on quartz or sapphire with high density (up to 100 CNTs/μm) or high semiconducting purity obtained through post-treatment (up to 99.9999%) (7–9), arrays with both high density and high purity have not yet been demonstrated. As a result, FETs built on CVD-grown CNT arrays suffer from either a low current on/off ratio (because of low purity) or a low driving current (because of low density) (2, 3). Also, each device based on CVD-grown

CNT arrays contains an undefined number of nanotubes, also known as an inconsistent pitch, which contributes to large device-to-device variation (2, 10, 11).

“Purified-and-placed” solution-processed CNT materials can provide CNTs with high semiconducting purity. This approach is simple and scalable, and it could provide wafer-scale assembly capability (12–18), but challenges remain. The current level of semiconducting purity of 99.99% must be improved to 99.9999% by further sorting CNTs and upgrading the purity characterization method. The solution-derived CNTs must be aligned into arrays with a 5- to 10-nm consistent pitch and full wafer coverage (2, 3, 5, 18, 19). A variety of methods have been developed to assemble solution-processed CNTs into arrays on substrates (20–25).

Arnold and colleagues (20, 21) reported a method to produce CNT arrays with a density of 47 CNTs/μm and built quasi-ballistic FETs with a high on-state current. However, the FETs thus fabricated suffered from low transconductance (0.1 mS/μm) because of low CNT density and poor gate efficiency, and the CNT arrays used had a strip-like shape that could not fully cover the substrate. Additionally, high-density and full-coverage CNT arrays have been prepared through methods such as the Langmuir-Schaefer-based method (23) or the vacuum filtration method (26). However, the CNT arrays thus obtained typically have a very high CNT array density of >400 CNTs/μm, resulting in a low on-state performance (in terms of transconductance g_m and on-state current I_{on} per tube) and a low current on/off ratio because of inefficient metal contacts with CNTs as well as serious screening effects from deleterious intertube interactions (3, 21, 27). To

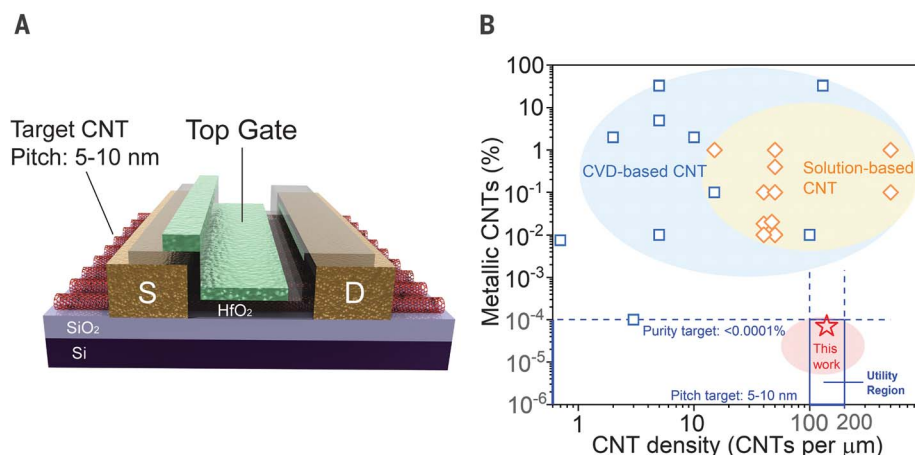


Fig. 1. Transistor structure and material target for CNT FET-based digital IC technology. (A) Schematic diagram showing a CNT-based top-gate FET with an ideal 5- to 10-nm CNT pitch. S, source; D, drain. (B) Semiconducting purity versus density of CNT arrays. The utility region is marked as a blue hollow box, and our results are located in the pink region, with a typical one marked as a red star. The extracted data (blue rectangles and orange diamonds) are listed in table S1 (7–9, 17, 20–23, 29, 32–44).

¹Key Laboratory for the Physics and Chemistry of Nanodevices and Center for Carbon-based Electronics, Department of Electronics, Peking University, Beijing 100871, China. ²Hunan Institute of Advanced Sensing and Information Technology, Xiangtan University, Hunan 411105, China. ³State Key Laboratory of Silicon Materials, School of Materials Science and Engineering, Zhejiang University, Hangzhou, 310027, China. ⁴Frontiers Science Center for Nano-optoelectronics, Peking University, Beijing 100871, China.

*These authors contributed equally to this work.

†Corresponding author. Email: zyzhang@pku.edu.cn (Z.Z.); lmpeng@pku.edu.cn (L.-M.P.)

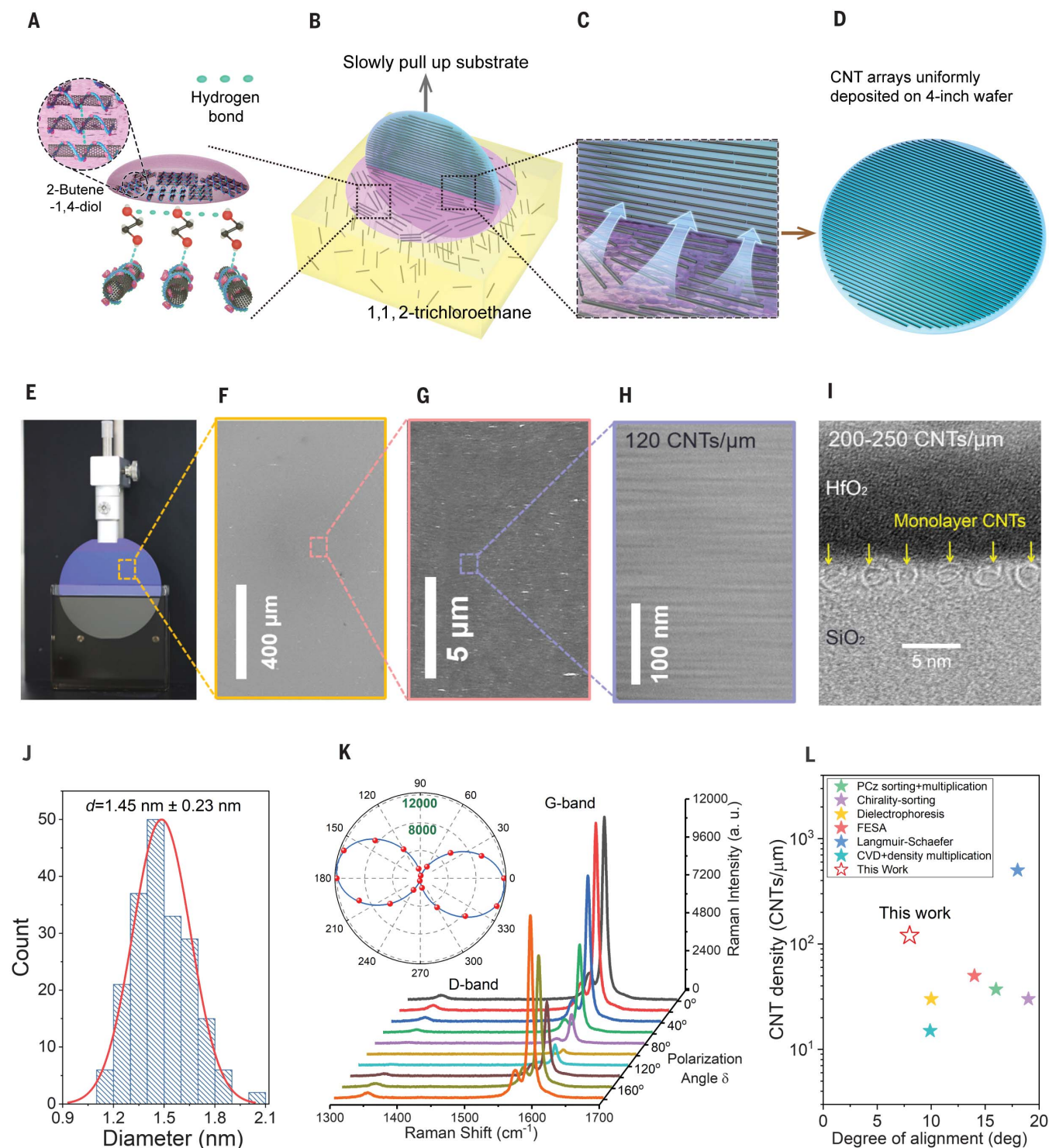
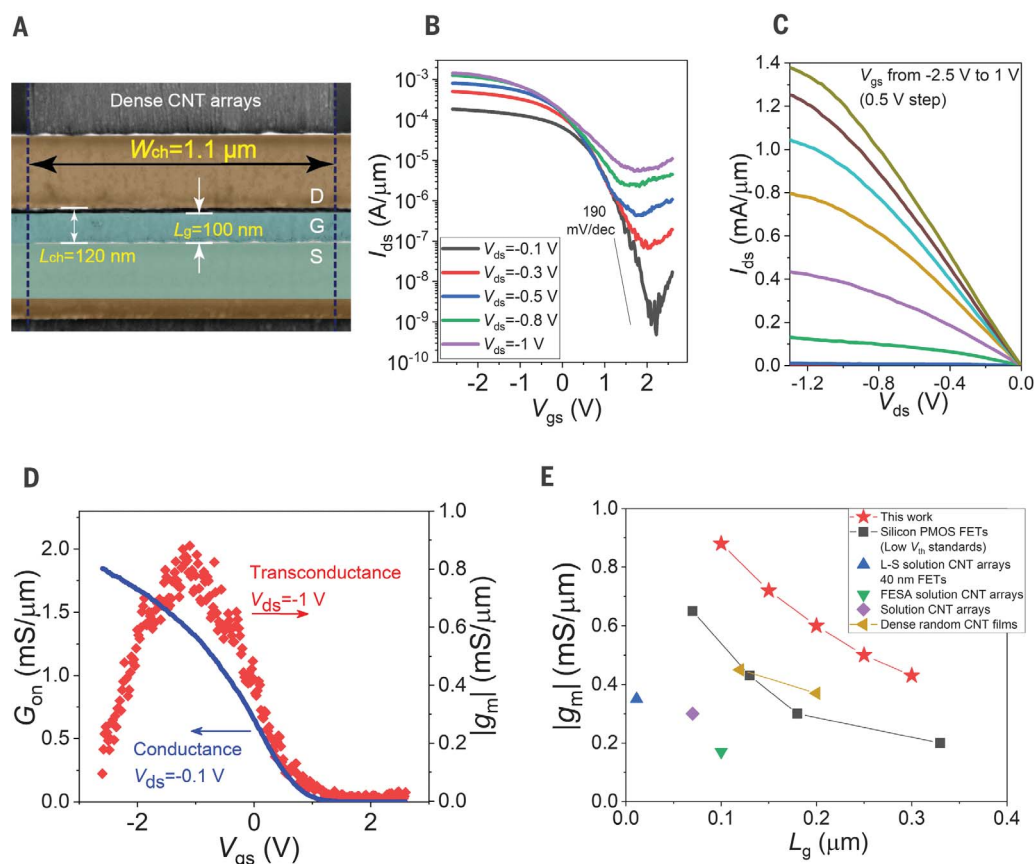


Fig. 2. Preparation and characterization of an A-CNT array. (A to D) Schematic images showing the process of preparing a wafer-scale A-CNT array. (E) Optical image showing the dip-coating setup used for coating CNTs on a 4-inch silicon wafer. (F to H) SEM images showing an as-deposited CNT array obtained using an optimal CNT solution concentration (40 $\mu\text{g/ml}$) at different magnifications. (I) Cross-section TEM image of a CNT array obtained using the highest CNT solution concentration of 60 $\mu\text{g/ml}$. The CNTs clearly show a density of at least 200 CNTs/ μm (<5 nm pitch); even at this high CNT density, our CNT arrays still

present a good monolayer property, which is crucial for electronics applications. (J) Diameter distribution of 200 CNTs measured by TEM. (K) Polarized Raman spectra of CNTs for different incident angles δ . The polarization angle of the incident light was changed through rotating a $\lambda/2$ wave plate during the measurements, and the reference (zero angle) is not exactly parallel to the CNT array. Inset: The Raman intensity is extracted and fitted in polar coordinates. (L) Benchmarking of the degree of alignment as a function of CNT density. All data used here are listed in table S2 (9, 20, 23, 25, 49, 50).

Fig. 3. Characteristics and benchmarking of A-CNT array-based top-gate FETs. (A) SEM image with a detailed channel region and top-view structure of a typical CNT FET. (B and C) Transfer characteristics (B) and output characteristics (C) of the CNT FET. (D) Low-bias ($V_{ds} = -0.1$ V) linear region conductance G_{on} and peak transconductance g_m in the saturated region ($V_{ds} = -1$ V), showing very high on-state performance of the device. (E) Benchmarking of g_m versus L_g of our results with reported champion CNT FETs and conventional commercial Si PMOS transistors (15, 21, 29, 31, 32, 53–55).



date, as the most important performance metric for FETs (28), the value of g_m in all fabricated FETs based on CNT arrays is usually <0.4 mS/ μ m (29), versus 0.5 mS/ μ m for Si CMOS FETs at a similar characteristic length. As a result, the stage delay of A-CNT ICs (consisting of FETs with a 100-nm channel length) measured with a ring oscillator (RO) was ~ 355 ps (30), which is at least one order of magnitude slower than that of similar silicon ICs or randomly oriented CNT film ICs of similar size (15, 16, 31).

Here, we report a multiple-dispersion sorting process to achieve a solution containing CNTs with a diameter distribution of 1.45 ± 0.23 nm and a semiconducting purity of $>99.9999\%$ according to a series of spectroscopy characterizations (fig. S1) and electrical measurements of 1300 FETs containing at least 2 million CNTs (figs. S2 to S4). A dimension-limited self-alignment (DLSA) procedure was developed to prepare well-aligned CNT arrays on a 4-inch (10 cm) wafer with a tunable density ranging from 100 to 200 CNTs/ μ m, which meets the fundamental requirements for CNTs to be useful for large-scale (but not industrial) IC fabrication as shown in Fig. 1B (7–9, 17, 20–23, 29, 32–44). The FETs and ICs based on the DLSA-processed A-CNT arrays with optimized structure and process exhibit real performance exceeding that of conventional Si CMOS transistors.

Semiconducting CNT arrays with ultrahigh purity and tunable density

In our multiple-dispersion sorting process for preparing CNTs of ultrahigh semiconducting purity (fig. S1) (45), raw CNTs are dispersed and sorted in toluene solvent by using conjugated PCz (poly[9-(1-octyloxy)-9H-carbazole-2,7-diyl]) molecules (shown schematically in fig. S1A) (46). This method was previously demonstrated to provide high selectivity for semiconducting CNTs (see the absorbance spectrum of PCz-sorted CNTs in toluene solvent in fig. S1D), with diameters narrowly distributed around 1.5 nm (46). After washing with tetrahydrofuran, the PCz-wrapped semiconducting CNTs were then redispersed in 1,1,2-trichloroethane (fig. S1C), and these processes were repeated twice. The second and third dispersion processes were crucial for removing excess PCz molecules to provide high electrical quality and monodispersed CNTs and to prevent the formation of CNT aggregates in solution, which was important for the subsequent alignment of CNTs into arrays. During each of the repeated dispersion processes, semiconducting CNTs could be further selected and purified (see fig. S1D for the absorbance spectra of redispersed CNTs in the target solvent) to achieve an extremely high semiconducting purity of $>99.9999\%$ (after multiple dispersion) according to statistical

electrical characterization of a total of 1300 wide-channel A-CNT FETs (figs. S2 to S4) (45). The purity of the resulting CNTs could in principle be further improved by increasing the number of dispersion processes.

We developed a DLSA procedure to assemble A-CNT arrays on a 4-inch wafer (Fig. 2, A to D). After a wafer was vertically inserted in the CNT solution, a thin layer was formed on the top surface by dropping 40 μ l of 2-buten-1,4-diol ($C_4H_8O_2$) close to the wafer, and this layer quickly spread around the wafer [see (45) for selection criteria of the top layer ($C_4H_8O_2$) and more details on DLSA]. The possible formation of hydrogen bonds (Fig. 2A) between PCz molecules (N atoms) and $C_4H_8O_2$ (H atoms in hydroxyl) allowed the PCz-wrapped CNTs with three-dimensional random orientations in the lower solvent to randomly walk into the surface region and become confined on the two-dimensional interface between the $C_4H_8O_2$ and the 1,1,2-trichloroethane solution. As the wafer was slowly pulled out (Fig. 2C), those CNTs confined on the interface then assembled onto the wafer surface through the strong affinity of $C_4H_8O_2$ and SiO_2 . CNTs were self-assembled along the contact line (horizontal orientation) between the wafer and interface by dimension-limited rotational degrees of freedom (20, 22), hence the name DLSA for this process.

The CNTs in the 1,1,2-trichloroethane solution did not adhere to the surface of the Si/SiO₂ wafer (fig. S5), which is the precondition for the DLSA method to work. Unlike the previously reported floating evaporative self-assembly method (20, 21), the DLSA method provided full CNT coverage across the entire substrate (Fig. 2D) because the CNTs in the lower 1,1,2-trichloroethane solvent were of sufficient quantity to serve as a continuous supply of CNTs to the interface of the binary liquid system. This solution-based DLSA procedure can be maintained in quasi-dynamic equilibrium, and the substrate withdrawal speed can thus be customized at a suitable range that depends on the interface absorption rate of CNTs. The density of the CNT array can thus be controlled through the CNT concentration in the 1,1,2-trichloroethane solution (fig. S6).

An optical image of the dip-coating mechanical apparatus for preparing A-CNT arrays on a 4-inch Si wafer with the DLSA method is shown in Fig. 2E, with the CNT coverage region marked in blue. The zoomed-in scanning electron microscopy (SEM) images (Fig. 2, F to H) of the CNT arrays show details of these arrays, in particular an excellent uniformity across hundreds of micrometers (figs. S7 to S9). The high-resolution SEM image (Fig. 2H) shows a typical CNT array (using a CNT solution concentration of 40 $\mu\text{g}/\text{ml}$) covered by a 10-nm HfO₂ thin film [grown by atomic-layer deposition (ALD)], with an optimal CNT density of ~ 120 CNTs/ μm , or tube-to-tube separation of 8 nm (fig. S6).

The cross-sectional high-resolution transmission electron microscopy (TEM) image (Fig. 2I) revealed that the prepared CNT array remained as a monolayer, even when the array density was increased to 200 to 250 CNTs/ μm (by using the highest CNT solution concentration, 60 $\mu\text{g}/\text{ml}$, in 1,1,2-trichloroethane). Monolayer formation is crucial to ensure excellent electrostatic properties for CNT-array FET applications [atomic force microscope (AFM) characterization in fig. S10 further verified the monolayer property of the CNT arrays]. Extensive characterizations confirmed that the DLSA method can produce the required monolayer CNT array with a suitable density for high-performance electronics applications, which is predicted to be in the range of 100 to 200 CNTs/ μm (corresponding to a CNT spacing of 5 to 10 nm) (2, 3). Detailed TEM examination of hundreds of CNTs revealed that the CNTs in the devices had a narrow diameter distribution of 1.45 ± 0.23 nm (Fig. 2J, measured by TEM), which lies well within the diameter requirement of 1.2 to 1.7 nm for realizing good ohmic contacts with relevant n- and p-type contact metals (2, 47, 48).

The polarized Raman spectra of CNTs for different incident angles δ (Fig. 2K) through

rotating the optics and the Raman intensity plot in polar coordinates (Fig. 2K, inset) indicate a large intensity ratio of 45 between the maximum Raman intensity (I_{max}) and minimum Raman intensity (I_{min}). This finding shows that the alignment between CNTs in the array was excellent at 8.3° [see (45) for calculation details; see fig. S11 for more information on the alignment uniformity across the wafer]. According to the benchmarking of the degree of alignment (Fig. 2L), the DLSA-prepared CNT arrays showed a narrower angular distribution and a higher CNT density than other reported CNT arrays produced by different methods (9, 20, 23, 25, 49, 50).

Top-gate CNT-array FETs with performance exceeding that of silicon FETs

Top-gate FETs were fabricated to explore the potential of the DLSA-prepared CNT arrays for electronics applications. Figure 3A shows a SEM image of a typical top-gate CNT FET [see (45)

and fig. S12 for detailed fabrication process flow]. Unlike the fabrication of usual CNT thin-film FETs, cleaning the as-produced CNT arrays before device fabrication was important for DLSA-prepared CNT-array FETs. In particular, our processes included a 600°C annealing process and a yttrium oxide-based coating and decoating process (15, 16, 51). In addition, we used an asymmetrical partial gate (with a gate length as short as 100 nm; see Figs. 1A and 3A) structure to improve the current on/off ratio at high voltage bias of -1 V (52). The height fluctuation in our monolayer CNT array channel was small (an AFM height profile is shown in fig. S10), so an ALD-grown HfO₂ gate insulator was thinned down to 7.3 nm (with a dielectric constant of ~ 16.8) to provide a high gate efficiency.

Our DLSA-prepared CNT-array FETs exhibited an on-state current exceeding 1.3 mA/ μm under a bias voltage of -1 V and a low bias ($V_{\text{ds}} = -0.1$ V) current on/off ratio greater than

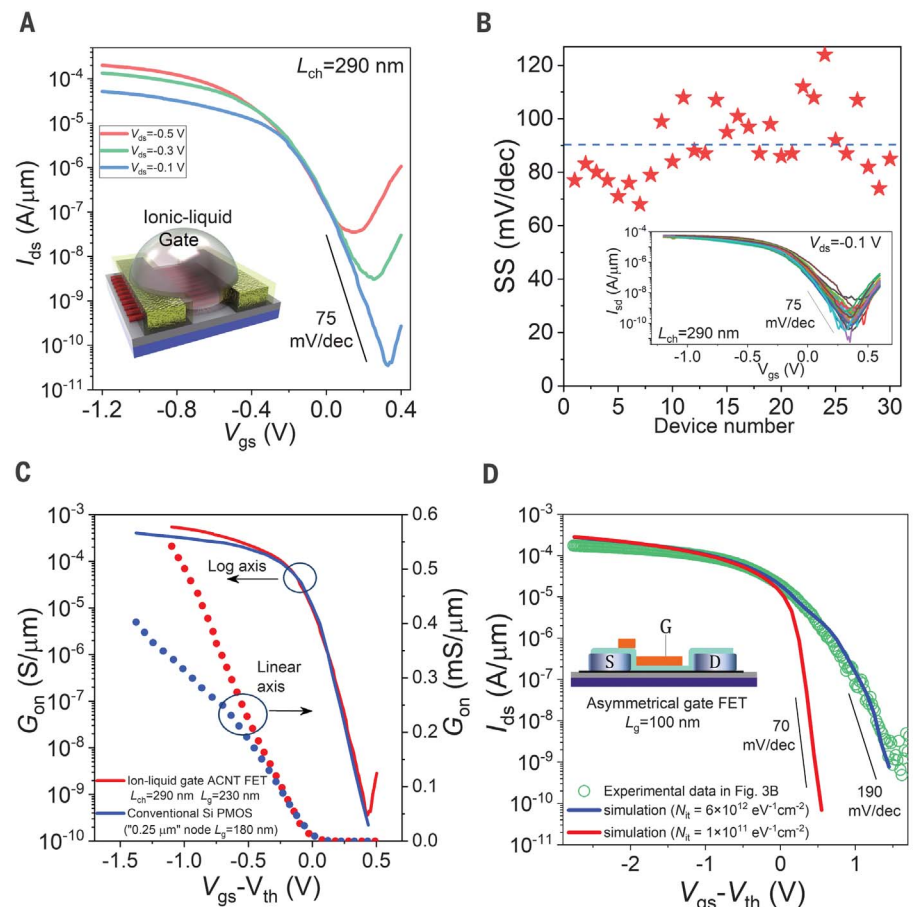
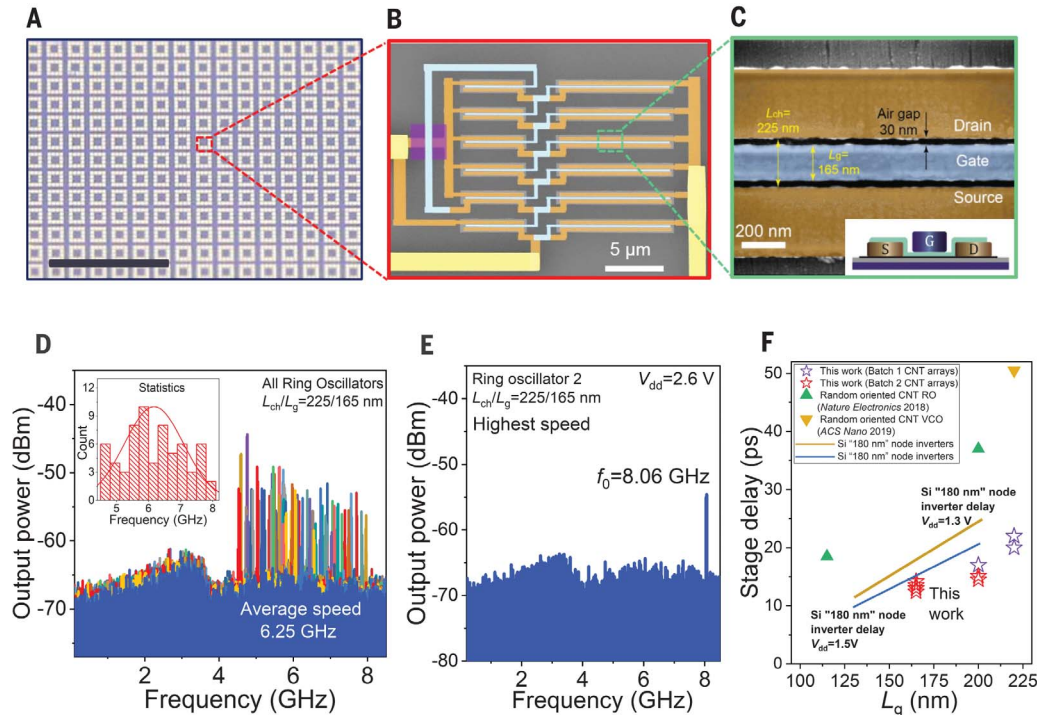


Fig. 4. Ionic-liquid gate CNT-array FETs. (A) Transfer characteristics of a typical FET with $L_{\text{ch}} = 290$ nm. Inset: Structure diagram of the ionic-liquid FET. (B) SS distribution of 30 ionic-liquid gate devices. Inset: Transfer characteristics of all 30 FETs. (C) Direct comparison of the transfer characteristics between a CNT-array FET and a Si high-performance standard PMOS FET (54). (D) Theoretically predicted transfer characteristics of CNT-array FETs with different interface state densities compared with the experimental results obtained from a top-gate CNT FET (corresponding to Fig. 3B). Inset: Device structure for both experiment and simulation.

Fig. 5. Structure and characteristics of CNT five-stage ROs.

(A) Optical image of batch-fabricated CNT five-stage ROs. Scale bar, 1.5 mm. (B and C) False-colored SEM images of a RO. The inset of (C) shows the gate structure of the CNT FET used for constructing the RO. (D) Power spectra of 65 CNT ROs under $V_{dd} = 2.5$ V; the inset shows statistical results of the switching frequency. (E) Power spectrum from the champion RO with the highest stage switching speed of 80.6 GHz. (F) Benchmarking of the stage delay of our champion ROs with state-of-the-art “0.18 μm ” silicon inverters and other champion CNT ROs with similar gate lengths (15, 16, 31).



10^5 (Fig. 3, B and C). The peak transconductance g_m of the FET reached 0.9 mS/ μm (at $V_{ds} = -1$ V; Fig. 3D), which is much higher than that of all reported CNT-based high-performance FETs (15, 16, 18, 21, 23–25, 29, 30, 32). Considering that the CNT density was ~ 100 CNTs/ μm , the peak g_m was converted to ~ 9 μS per CNT, which is the highest among all CNT film transistors (15, 16, 18, 21, 23–25, 29, 30, 32). This value suggested that the outstanding performance of the CNT-array FET originated from contributions of all CNTs in the array, even for such a high array density. These typical CNT-array FETs exhibited higher g_m than that of Si FETs with similar gate length, including 0.13-, 0.18-, 0.25-, and 0.35- μm nodes (high-performance standards; Fig. 3E) (31, 53–55). When benchmarked against other reported high-performance CNT-based FETs (with current on/off ratio at least two orders of magnitude) (15, 21, 29, 32), our device showed not only much higher g_m but also lower subthreshold swing (SS) (see fig. S13). The simultaneous high g_m and low SS resulted from excellent CNT-array films with high density, good alignment, and high semiconducting purity, as well as an optimized device fabrication procedure that leads to very clean materials, excellent contacts, and a high gate efficiency for every CNT in the channel.

The SS of the top-gate CNT-array FET ranged from 100 to 200 mV/decade, which is better than that of most high-performance CNT-based FETs. However, this value still falls below the standard SS requirement (below 100 mV/decade) for digital ICs and is much

higher than that based on individual CNTs with similar gate efficiency (5). One important factor that contributes to the SS degradation is the diameter variations of the CNTs in the arrays used in the FET channels. By a simple theoretical analysis using the virtual source model and Monte Carlo method, we found that the variation arising from the diameter distribution of our CNTs (1.45 ± 0.23 nm) would only degrade SS down to ~ 65 mV/decade for the top-gate CNT-array FETs (fig. S14). However, the interface-trapped charge density around the CNT channel degraded SS because these trapped charges may severely screen the electric field and lower the gate efficiency.

The effect of the interface-trapped charges on the SS of a typical FET can be discussed using the formula

$$\begin{aligned} \text{SS} &= \frac{dV_g}{d(\log I_{ds})} \\ &= 2.3 \frac{mkT}{q} \approx 2.3 \left(1 + \frac{qN_{it}}{C_g} \right) \frac{kT}{q} \quad (1) \end{aligned}$$

(56), where q is the elementary charge, k is the Boltzmann constant, T is temperature, and m is referred to as the ideality factor, which is determined mainly by the interface state fixed charge density (N_{it}) and gate capacitance C_g . N_{it} is well established to be on the order of 10^{12} eV $^{-1}$ cm $^{-2}$ in solution-derived CNT film FETs (56), which is two orders of magnitude higher than that in conventional Si CMOS FETs (57).

This large charge density N_{it} contributes to the nonideal subthreshold performance of the CNT-array FET or large SS. Lowering N_{it} during device fabrication is difficult because it mainly results from the polymer residues wrapping the CNTs. The most effective way to improve SS would be to further improve the gate efficiency (i.e., increase C_g). We thus constructed ionic liquid (IL)-gated FETs based on the DLSA-prepared CNT arrays (Fig. 4), where the electric double layers at the IL/solid interface act as nanogap capacitors with extremely large capacitance (58–60) [see (45) and fig. S15 for the fabrication and measurement setup].

The adoption of an ultrahigh-efficiency IL gate improved the switching-off property of CNT-array FETs. In particular, it lowered the SS of a typical CNT-array FET to 75 mV/decade (Fig. 4A). The SS values of 30 IL-gate devices were distributed in a narrow range, with an average value of ~ 90 mV/decade (Fig. 4B), and the SS values of some devices even approached the 60 mV/decade theoretical limit at room temperature. A direct comparison of the transfer characteristics of an IL-gated CNT-array FET ($L_{ch} = 290$ nm) and those of a commercial Si PMOS (p-type metal-oxide semiconductor) FET with similar gate length (0.25- μm node with physical gate length of 0.18 μm ; Fig. 4C) (54) showed that the CNT-array FET exhibited better on-state current and similar off-state current in a smaller V_{gs} range than that of the Si PMOS FET.

Although the IL gate is not suitable for scalable integration of solid-state devices, it

reflects the potential of CNT-array FETs with enhanced gate efficiency or lowered interface state density. Quantitative simulations revealed that if the interface state density could be lowered to $10^{11} \text{ eV}^{-1} \text{ cm}^{-2}$ (still higher than that in Si CMOS FETs), then the CNT-array FET with a solid-state gate dielectric could improve SS from 190 to 70 mV/decade (Fig. 4D and table S3) (45).

CNT array-based ring oscillators with a frequency of >8 GHz

Relative to a randomly oriented CNT film, an aligned CNT array with high purity and suitable density should provide better circuit performance due to notably enhanced current driving ability as well as smaller intrinsic capacitance. To demonstrate this, we used our DLSA-prepared, wafer-scale CNT arrays to construct high-performance ICs, particularly the RO circuit, which is a special standard test IC for assessing the performance and uniformity of new IC technology. We fabricated hundreds of top-gate five-stage ROs in a $5 \text{ mm} \times 5 \text{ mm}$ region [see the optical image and SEM images in Fig. 5, A to C; see (45) and fig. S16 for the fabrication process flow] to directly test the stage propagation delay of inverters by characterizing the actual switching frequency of ROs. The top-gate structure used here was optimized to reduce parasitic capacitances between the gate and source/drain (with a 30-nm air gap on each side of the gate) and to reduce the gate resistance (tall metal gate) (15, 16).

The channel and gate lengths of the CNT-array FETs used for constructing ROs were designed to be 225 and 165 nm, respectively. Typical output and transfer characteristics of the FETs are shown in fig. S17, with an on-state current of $\sim 0.75 \text{ mA}/\mu\text{m}$ and a peak g_m of $>0.5 \text{ mS}/\mu\text{m}$. We measured 128 five-stage ROs in one region of the wafer (Fig. 5A), among which 65 ROs functioned successfully, indicating a yield of $>50\%$, which is a relatively high yield among laboratory-fabricated ROs. (15, 30) The frequency spectra of these ROs are shown in Fig. 5D; the corresponding oscillating frequency f_o ranged from 4.7 to 8 GHz, with an average RO switching frequency of $f_o = 6.25 \text{ GHz}$ under supply voltage $V_{dd} = 2.5 \text{ V}$. The ROs also oscillated well ($f_o = 7 \text{ GHz}$) under much lower V_{dd} (down to 1.8 V; see fig. S18). The highest f_o reached 8.06 GHz at $V_{dd} = 2.6 \text{ V}$ (Fig. 5E), corresponding to a stage-switching speed of 80.6 GHz and a stage delay of 12.4 ps.

We benchmarked these results with the actual speed of several representative types of IC technologies based on the measured stage delay according to different benchmarking conditions (V_{dd} or gate length L_g ; Fig. 5F and fig. S18B). The DLSA-prepared CNT array-based ROs displayed lower gate delays than

all reported nanomaterials-based ROs with similar gate lengths and under lower V_{dd} . In addition, our CNT array-based ICs exhibited real performance (speed) exceeding that of conventional Si CMOS ICs under similar gate lengths (Fig. 5F) (31).

Outlook

We showed, by combining multiple-dispersion sorting and DLSA methods, that well-aligned (within 9° of alignment), high-purity (better than 99.9999%), and high-density (tunable between 100 and 200 CNTs/ μm) arrays of CNTs can be prepared on 4-inch silicon wafers with full wafer coverage; these CNT arrays meet the fundamental requirements for large-scale fabrication of digital ICs. Preliminary demonstrations using DLSA-prepared CNT arrays show that these CNT-array FETs and ICs outperform those of silicon technology with similar characteristic lengths in several key performance metrics.

Further development of this CNT-based platform will require optimization of both the material preparation and corresponding device fabrication processes. First, further improvement of the uniformity of the tube-to-tube pitch, direction, and diameter of CNTs on a large scale (such as on an 8-inch wafer) is necessary for ultralarge-scale integration of CNT ICs, particularly for sub-10 nm technology nodes. Moreover, the CNTs in the array need to be further cleaned. A certain amount of polymer residue remains wrapped around the CNTs; this prevents the formation of better contacts with smaller resistance at the source/drain and contributes to the high interface charge density (N_{it}) in the gate stack of CNT FETs (15, 16, 51). Decoupling the polymer residues from the CNT arrays while not introducing additional damage is an important issue for the fabrication of high-performance, high-reliability transistors using DLSA-prepared CNT arrays. The adoption of the multilayer interconnect technology widely used in Si technology and the optimization of the device structure would also be expected to further improve the working speed of CNT-based ICs.

REFERENCES AND NOTES

- Semiconductor Industry Association, 2015 *International Technology Roadmap for Semiconductors* (2015); www.semiconductors.org/resources/2015-international-technology-roadmap-for-semiconductors-itsr/.
- G. S. Tulevski et al., *ACS Nano* **8**, 8730–8745 (2014).
- A. D. Franklin, *Nature* **498**, 443–444 (2013).
- R. Chau et al., *IEEE Trans. NanoTechnol.* **4**, 153–158 (2005).
- C. Qiu et al., *Science* **355**, 271–276 (2017).
- P. Zhang et al., *Nano Res.* **8**, 1005–1016 (2015).
- Y. Hu et al., *Nat. Commun.* **6**, 6099 (2015).
- S. H. Jin et al., *Nat. Nanotechnol.* **8**, 347–355 (2013).
- J. Si et al., *ACS Nano* **12**, 627–634 (2018).
- N. Patil et al., *IEEE Trans. NanoTechnol.* **10**, 744–750 (2010).
- J. Zhang et al., *IEEE Trans. Comput. Aided Des. Integrated Circ. Syst.* **31**, 453–471 (2012).
- C. Qiu et al., *Adv. Mater.* **31**, e1800750 (2019).

- B. Chen et al., *Nano Lett.* **16**, 5120–5128 (2016).
- Y. Yang, L. Ding, J. Han, Z. Zhang, L.-M. Peng, *ACS Nano* **11**, 4124–4132 (2017).
- D. Zhong et al., *Nat. Electron.* **1**, 40–45 (2018).
- L. Liu et al., *ACS Nano* **13**, 2526–2535 (2019).
- G. Hills et al., *Nature* **572**, 595–602 (2019).
- C. Zhao et al., *Adv. Funct. Mater.* **29**, 1808574 (2019).
- C.-S. Lee, E. Pop, A. D. Franklin, W. Haensch, H.-S. P. Wong, *IEEE Trans. Electron Dev.* **62**, 3070–3078 (2015).
- Y. Joo, G. J. Brady, M. S. Arnold, P. Gopalan, *Langmuir* **30**, 3460–3466 (2014).
- G. J. Brady et al., *Sci. Adv.* **2**, e1601240 (2016).
- M. Engel et al., *ACS Nano* **2**, 2445–2452 (2008).
- Q. Cao et al., *Nat. Nanotechnol.* **8**, 180–186 (2013).
- T. A. Shastri et al., *Small* **9**, 45–51 (2013).
- M. G. Zhu, J. Si, Z. Zhang, L. M. Peng, *Adv. Mater.* **30**, e1707068 (2018).
- X. He et al., *Nat. Nanotechnol.* **11**, 633–638 (2016).
- F. Léonard, *Nanotechnology* **17**, 2381–2385 (2006).
- S. M. Sze, K. K. Ng, *Physics of Semiconductor Devices* (Wiley, 2006).
- Q. Cao, J. Tersoff, D. B. Farmer, Y. Zhu, S.-J. Han, *Science* **356**, 1369–1372 (2017).
- S.-J. Han et al., *Nat. Nanotechnol.* **12**, 861–865 (2017).
- S. Yang et al., in *International Electron Devices Meeting 1998. Technical Digest* (IEEE, 1998), pp. 197–200.
- Y. Cao et al., *ACS Nano* **10**, 6782–6790 (2016).
- L. Ding et al., *Nano Lett.* **9**, 800–805 (2009).
- W. Zhou, S. Zhan, L. Ding, J. Liu, *J. Am. Chem. Soc.* **134**, 14019–14026 (2012).
- M. M. Shulaker et al., *Nature* **501**, 526–530 (2013).
- F. Yang et al., *ACS Nano* **11**, 186–193 (2017).
- X. Xie et al., *Nat. Commun.* **5**, 5332 (2014).
- M. M. Shulaker et al., in *2014 IEEE International Electron Devices Meeting* (IEEE, 2014), pp. 33.36.31–33.36.34.
- C. Kocabas et al., *Proc. Natl. Acad. Sci. U.S.A.* **105**, 1405–1409 (2008).
- G. J. Brady et al., *ACS Nano* **8**, 11614–11621 (2014).
- Q. Cao, S. J. Han, G. S. Tulevski, *Nat. Commun.* **5**, 5071 (2014).
- G. J. Brady, Y. Joo, S. Singha Roy, P. Gopalan, M. S. Arnold, *Appl. Phys. Lett.* **104**, 083107 (2014).
- M. Steiner et al., *Appl. Phys. Lett.* **101**, 053123 (2012).
- Y. Cao, Y. Che, H. Gui, X. Cao, C. Zhou, *Nano Res.* **9**, 363–371 (2016).
- See supplementary materials.
- J. Gu et al., *Small* **12**, 4993–4999 (2016).
- W. Kim et al., *Appl. Phys. Lett.* **87**, 173101 (2005).
- Z. Chen, J. Appenzeller, J. Knoch, Y.-M. Lin, P. Avouris, *Nano Lett.* **5**, 1497–1502 (2005).
- Z. Ma et al., *Appl. Phys. Lett.* **108**, 063114 (2016).
- S. Shekhar, P. Stokes, S. I. Khondaker, *ACS Nano* **5**, 1739–1746 (2011).
- Z. Ma, J. Han, S. Yao, S. Wang, L.-M. Peng, *ACS Appl. Mater. Interfaces* **11**, 11736–11742 (2019).
- C. Qiu et al., *ACS Nano* **9**, 969–977 (2015).
- S. Tyagi et al., in *International Electron Devices Meeting 2000. Technical Digest* (IEEE, 2000), pp. 567–570.
- A. Brand et al., *Intel Tech. J. Q.* **3**, 1998 (1998).
- M. Bohr et al., in *Proceedings of 1994 IEEE International Electron Devices Meeting* (IEEE, 1994), pp. 273–276.
- L. Xu, N. Gao, Z. Zhang, L.-M. Peng, *Appl. Phys. Lett.* **113**, 083105 (2018).
- G. Kapila, B. Kaczar, A. Nackaerts, N. Collaert, G. Groeseneken, *IEEE Electron Device Lett.* **28**, 232–234 (2007).
- M. M. Perera et al., *ACS Nano* **7**, 4449–4458 (2013).
- J. Ye et al., *Proc. Natl. Acad. Sci. U.S.A.* **108**, 13002–13006 (2011).
- J. Xia, F. Chen, J. Li, N. Tao, *Nat. Nanotechnol.* **4**, 505–509 (2009).

ACKNOWLEDGMENTS

We thank D. Zhong (Stanford University) for useful discussion, W. Liu (Peking University) for AFM technique support, and J. Yang (Peking University) for discussions on polarized Raman data. **Funding:** Supported by the National Key Research & Development Program (grants 2016YFA0201901 and 2016YFA0201902), the National Natural Science Foundation of China (grant 61888102), and the Beijing Municipal Science and Technology Commission (grant Z181100004418011). **Author contributions:** Z.Z. and L.-M.P. were in charge of and advised and led on this project; L.L. was involved in most aspects of this work, from CNT material

characterization to device/IC fabrication and device/IC characterization, and analyzed the results with Z.Z. and L.-M.P.; J.H. designed and produced dense A-CNT arrays, including the multiple dispersion process as well as the DLSA process; L.L. and J.H. characterized the A-CNT material and worked together for material system improvement; L.L., H.S., and Z.M. performed the polarized Raman spectroscopy characterization; L.X. performed the SS/ N_{it} simulations based on top-gate devices using the

VS model, Monte Carlo model, and TCAD; M.X. helped improve the ion-liquid device measurement; J.Z., C.Z., and H.S. were involved in device fabrication; S.D. and C.J. performed TEM experiments; L.L., Z.Z., and L.-M.P. co-wrote the manuscript. All authors commented on and discussed this work. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6493/850/suppl/DC1
Materials and Methods
Figs. S1 to S18
Tables S1 to S3

16 December 2019; accepted 9 April 2020
10.1126/science.aba5980

REPORT

TOPOLOGICAL OPTICS

Observation of Floquet solitons in a topological bandgap

Sebabrata Mukherjee* and Mikael C. Rechtsman*

Topological protection is a universal phenomenon that applies to electronic, photonic, ultracold atomic, mechanical, and other systems. The vast majority of research in these systems has explored the linear domain, where interparticle interactions are negligible. We experimentally observed solitons—waves that propagate without changing shape as a result of nonlinearity—in a photonic Floquet topological insulator. These solitons exhibited distinct behavior in that they executed cyclotron-like orbits associated with the underlying topology. Specifically, we used a waveguide array with periodic variations along the waveguide axis, giving rise to nonzero winding number, and the nonlinearity arose from the optical Kerr effect. This result applies to a range of bosonic systems because it is described by the focusing nonlinear Schrödinger equation (equivalently, the attractive Gross-Pitaevskii equation).

The discovery of the integer quantum Hall effect (1) and its topological interpretation (2) initiated extensive research into exotic topological materials in a wide variety of platforms (3–14). The prediction of quantum Hall-like states for light (3) has led to wide interest (15, 16) in the interplay between topological protection and photonic properties, especially effects that are not realized for electrons in solids. After its first observation in a gyromagnetic photonic crystal at microwave frequencies (4), topological edge states were demonstrated at optical frequencies in waveguide lattices (5) and in ring resonators (6). The investigation into topological states in electromagnetic systems has been largely limited to the linear domain, where photons propagate independently, governed by Maxwell's equations. These topological states are described as a system of noninteracting particles with topologically nontrivial bands, characterized by integer-valued invariants such as Chern numbers.

Among the most fundamental effects in nonlinear optics is the Kerr effect: a variation

of the refractive index proportional to the local intensity of light. This intensity-dependent refractive index is a manifestation of the nonlinear dielectric polarization induced by optical fields. Thus, at high intensity, photons can effectively interact, mediated by the ambient medium. Indeed, the nonlinear Schrödinger equation describing the propagation of light through a nonlinear medium is equivalent to the Gross-Pitaevskii equation, which describes bosonic interactions in a Bose-Einstein condensate in the mean-field limit. Hence, photonic lattices are a natural platform for studying the interplay of topology and interparticle interactions.

Here, we observed optical spatial solitons (17–22) in an anomalous Floquet topological insulator (23–25), realized using a periodically modulated waveguide lattice. For such a topological phase, the Floquet driving gives rise to a nonzero winding number, implying the presence of topological edge modes (23); however, the standard topological invariants (e.g., Chern numbers) are zero (hence the name “anomalous”). A family of solitons spectrally

resides in the topological bandgap, and during propagation, the solitons execute cyclotron-like rotations inherited from the linear host lattice (we henceforth refer to a soliton in a topological bandgap as a “topological soliton”). Consistent with previous theoretical predictions of topological solitons, these solitons show behavior that arises from the topological nature of the system (26–29). Indeed, the cyclotron-like motion defines the topological character of the solitons: In the quantum Hall effect, the cyclotron motion gives rise to the “skipping orbits” that describe protected edge states (30). In that sense, these solitons are of a different nature from that of previously observed bandgap solitons.

In the presence of the optical Kerr effect, light propagation through a photonic lattice with nearest-neighbor evanescent coupling is described by the discrete nonlinear Schrödinger equation, under the paraxial approximation:

$$i \frac{\partial}{\partial z} \phi_s(z) = \sum_{(s')} H_{ss'} \phi_{s'} - |\phi_s|^2 \phi_s \quad (1)$$

where the propagation distance (z) plays the role of time ($z \leftrightarrow t$), and $H_{ss'}$ is the linear tight-binding Hamiltonian (the summation is over neighboring sites only). We define $|\phi_s|^2 = g|\psi_s|^2$ where $|\psi_s|^2$ is the optical power at the s th waveguide and g is determined by the nonlinear refractive index coefficient, the effective area of the waveguide modes, and the wavelength. At sufficiently low optical power, the nonlinear term of Eq. 1 is negligible. Here, we have used the self-focusing nonlinearity (corresponding to attractive interactions in the Gross-Pitaevskii equation), which was experimentally validated for the nonlinear medium used here (31). In the absence of optical losses, the total energy and the renormalized

Department of Physics, The Pennsylvania State University, University Park, PA 16802, USA.
*Corresponding author. Email: mukherjeesebabrata@gmail.com (S.M.); mcrworld@psu.edu (M.C.R.)

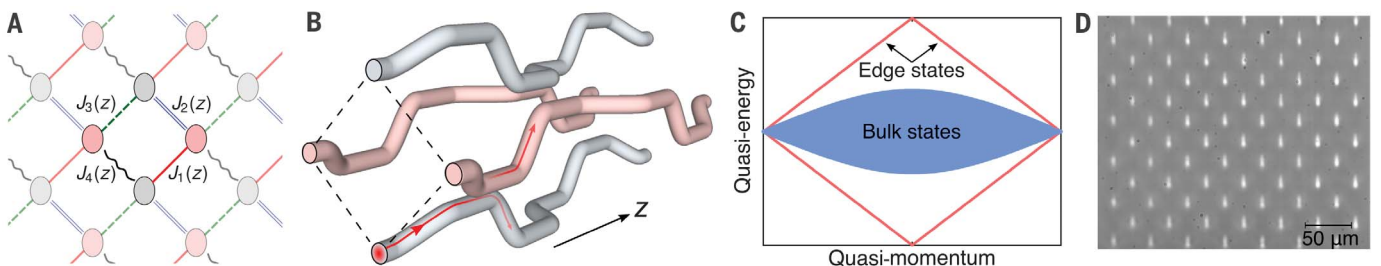


Fig. 1. Photonic implementation of an anomalous Floquet topological insulator. (A) A periodically driven square lattice where the four equal couplings $J_m(z)$ [$m = 1, \dots, 4$] are switched on and off in a cyclic (spatially and z periodic) manner. (B) Schematic showing how this driving protocol is implemented using three-dimensional waveguide arrays. Only four sites are shown here for one complete driving period, z_0 . (C) Quasi-energy spectrum in the linear regime (for the experimentally realized parameters) showing two ungapped bulk bands with zero net Chern number and chiral edge modes. (D) Micrograph of the facet of a driven photonic square lattice fabricated by femtosecond laser writing.

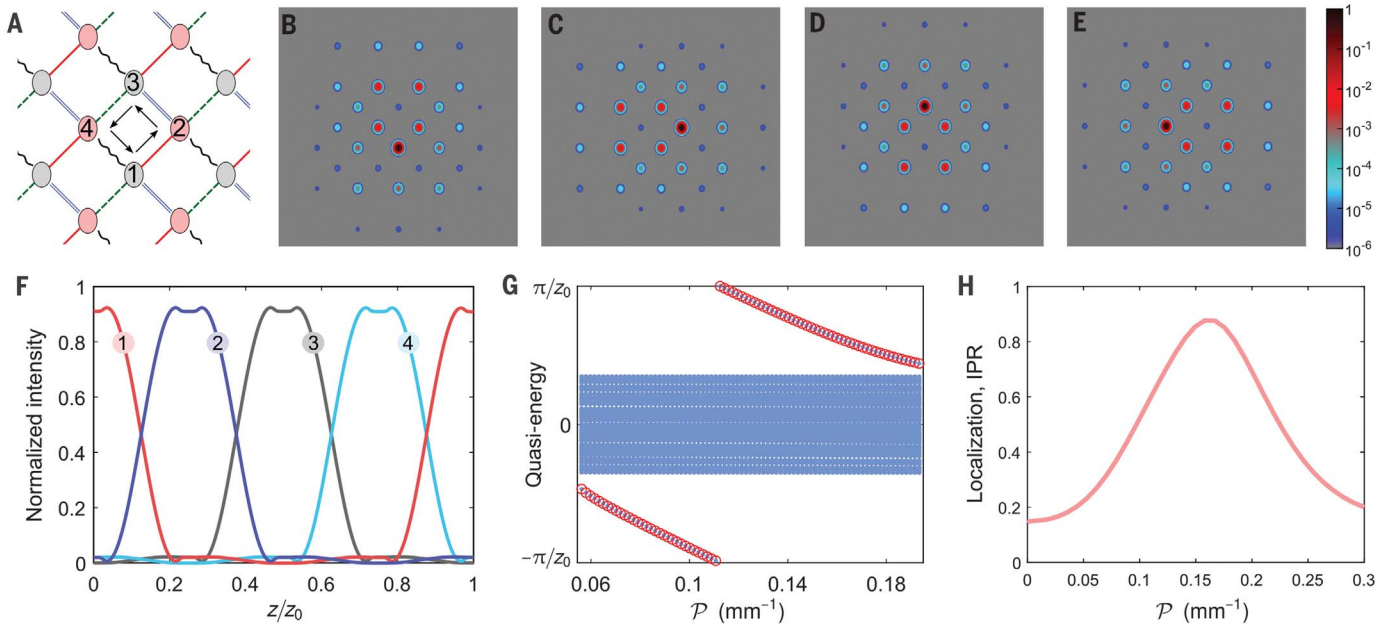


Fig. 2. Topological bandgap soliton performing cyclotron-like motion.

(A) Schematic showing the four sites (1 to 4) where the maximum optical power of a topological soliton is contained during propagation. (B to E) Normalized intensity profile of a soliton (at $P = 0.088 \text{ mm}^{-1}$) rotating counterclockwise and cycling back to itself after each complete period of driving; the color map is in log scale. Here, the soliton profile is shown for each quarter of a complete period—that is, $z = \{0, 1, 2, 3\}z_0/4$. (F) Variation of

normalized intensity at the four sites [1 to 4 in (A)] showing the dynamics in a complete period. (G) Quasi-energy as a function of renormalized power, showing the family of bulk solitons (red circles) on both sides of the linear modes. (H) A signature of topological solitons: When light is coupled to a single bulk waveguide, the output intensity pattern at $z = 2z_0$ exhibits a distinct feature: a peak in the inverse participation ratio (IPR), which is detected in experiments (see Fig. 3).

power ($P \equiv \sum_s |\phi_s|^2$) are conserved. Nonlinearity in the off-diagonal coupling term is negligible.

Consider a periodically modulated photonic square lattice (24, 25, 32) with nearest-neighbor couplings $J_m(z)$ [$m = 1, \dots, 4$] that are engineered in a cyclic (spatially and z periodic) manner such that every waveguide is coupled to only one of its nearest neighbors at a given propagation distance z (Fig. 1, A and B). The driving period z_0 is split into four equal steps, and over each quarter of the driving period, only one of the four couplings is switched on, with a fixed $\Lambda = \int dz J_m(z)$ (the integral is taken over one coupling operation). In the linear regime, the quasi-energy spectrum can be obtained by diagonalizing the Floquet evolution operator over one period, defined as

$$\hat{U}(z_0) = \mathcal{T} \exp \left[-i \int_0^{z_0} \hat{H}(\tilde{z}) d\tilde{z} \right] \quad (2)$$

where \mathcal{T} indicates the “time” ordering and $\hat{H}(z) = \hat{H}(z + z_0)$ is the periodic linear Hamiltonian. This driven lattice supports two ungapped bulk bands (henceforth referred to as the bulk band), and the bandwidth is determined by Λ : For $\Lambda = \pi/2$, the bulk band becomes perfectly flat, and the bandgap closes at $\Lambda = \{\pi/4, 3\pi/4\}$. To experimentally realize

a weakly dispersive bulk band with an appreciable bandgap, we set $\Lambda = 1.85 \pm 0.05$. The ratio of the bulk bandwidth to the maximal coupling strength $\max[J(z)]$ is estimated to be ~ 0.25 , which quantifies the flatness of the band (whereas the bandwidth of a standard static square lattice is eight times the coupling strength). Figure 1C shows a spectrum calculated for a strip geometry aligned along the vertical direction and periodic along the horizontal direction. As a result of the periodicity of quasi-energy, the edge modes can cross the bandgap, connecting the top and bottom of the band structure. A single chiral edge mode exists above and below the bulk band (propagating in the same direction on a given edge), which implies that the Chern number of the bulk band is zero. For such anomalous Floquet topological insulators (23, 33), the topology can be captured using a different topological invariant, the winding number (31). This scenario can only arise in the presence of suitable time-periodic driving; anomalous Floquet topological insulators have no analog in static systems. There is only one bandgap in the system (Fig. 1C), and it is topological.

Using a self-consistency method modified for Floquet systems (26), we sought localized nonlinear solutions (solitons) in this modulated photonic square lattice (31). The result is solitons in the Floquet sense: Because of the z -periodic driving, the solitons reproduce

themselves after each complete period (up to a phase factor), although micromotion within the Floquet cycle is allowed for. The solitons continuously rotate, performing cyclotron-like motion (movie S1). Figure 2, B to E, shows the normalized intensity profile (i.e., $|\phi_s|^2/P$) of a soliton at each quarter-period (i.e., $z = \{0, 1, 2, 3\}z_0/4$). Figure 2F shows the variation of normalized intensity at the four sites (1 to 4 in Fig. 2A) where the maximum optical power of the soliton is supported during propagation.

The quasi-energy of a soliton is determined by the overall phase acquired after the propagation of one driving period ($\epsilon = \text{phase}/z_0$). The quasi-energy spectrum, plotted as a function of the renormalized power (Fig. 2G), shows a family of bandgap solitons (red circles) bifurcating from the linear band (blue). The size (i.e., the spatial extent) of the solitons first decreases as a function of power, showing maximal localization near the mid-gap quasi-energy π/z_0 . When the power is further increased, these Floquet solitons become delocalized as they approach the band from the other side (31)—a behavior unlike that of standard lattice solitons. In other words, for a given dispersion of the linear band, the spatial extent of these solitons is determined by their quasi-energy; solitons closer in energy to the linear band have a larger spatial extent (movie S2). Because the solitons are strongly

peaked on a single site, it is possible to probe them in experiments by means of single-site excitation (i.e., by coupling light into a single waveguide). A signature of these bandgap solitons can be experimentally detected by measuring the degree of localization of the output intensity patterns as a function of renormalized power. We plot the inverse participation ratio

$$\text{IPR} \equiv \frac{\sum |\phi_s|^4}{\left(\sum |\phi_s|^2\right)^2} \quad (3)$$

(a measure of localization), after two driving periods, in Fig. 2H. Here, we observed a clear peak in the IPR, corresponding to the existence and strong localization of these gap solitons. We note that the peak occurs at an input power higher than the power of the soliton at its most localized; this is because we do not input the exact soliton wave function but rather a single site, meaning that some power is lost to background radiation in the lattice. The trend in IPR (i.e., delocalization to localization to delocalization) is qualitatively different from the trend in a topologically trivial static lattice, where IPR continuously increases and then saturates at very high nonlinearity (31).

To demonstrate how topological solitons are distinct from trivial ones, we examined the area encircled by the center of mass of the soliton, which acts as a quantitative measure of whether the orbit can be considered “cyclotron-like.” We found that this area is finite for any soliton in a topologically non-trivial gap and is zero for any soliton in a topologically trivial gap; these findings cor-

respond directly to the cyclotron-like nature of the soliton micromotion in the topological case. As a stark example of this fact, we give an example of two different families of solitons that reside in the same lattice: one in a topological gap with nonzero encircled area, and one in a trivial gap with zero encircled area (31).

To experimentally probe the solitons described above, we coupled intense laser pulses into femtosecond laser-fabricated waveguide arrays [see (31) for fabrication details]. In this situation, ϕ_s is a function of both propagation distance and time t : $\phi_s = \phi_s(z, t)$. Because of the temporal shape of light pulses, self-phase modulation and chromatic dispersion are relevant. Laser pulses were temporally stretched (to $t_p \approx 2$ ps) and down-chirped such that these effects could be ignored (31). Additionally, we found that the insertion loss is independent of nonlinearity, implying negligible nonlinear loss due to multiphoton absorption. To validate these claims, we performed experiments with a topologically trivial static square lattice consisting of straight coupled waveguides. We observed that the output intensity pattern became increasingly localized as a function of input power, and finally, all the optical power was trapped largely in the single site where the light was launched at the input (31) (movie S3), as expected. This baseline experiment clearly demonstrates the formation of highly localized solitons in a topologically trivial bandgap (22, 34).

For the topological case, a 76-mm-long periodically modulated square lattice of 84 sites was fabricated with the previously mentioned driving parameters; a micrograph of this lattice (cross section) is shown in Fig. 1D. Initially,

the waveguides were separated by $26.5 \mu\text{m}$ such that the evanescent couplings were negligibly small. To couple any two desired waveguides, we first reduced the inter-waveguide separation by synchronously bending the waveguide paths, then kept the two waveguides parallel with $14.5\text{-}\mu\text{m}$ fixed center-to-center spacing, and finally separated them (Fig. 1B).

Nonlinear characterization of the photonic lattice is summarized in Fig. 3. For all measurements, we launched light pulses into a bulk waveguide away from the edges; during linear diffraction, the light did not reach the sides of the array, and thus any edge effects could be neglected. As detailed in (31), the renormalized power \mathcal{P} at the input of the lattice was found to be 0.076 mm^{-1} per unit average input power in mW (note that \mathcal{P} has the same dimension as the evanescent coupling strength, for clarity). In the first set of experiments, we measured output intensity distributions at $z = 2z_0$ as a function of average input power (movie S4). The variation of IPR with input power is shown in Fig. 3A. At low optical power (i.e., in the linear regime), this single-site excitation overlapped with the weakly dispersive bulk modes, and light diffracted away from the site into which it was injected (Fig. 3B). As input power was increased, output intensity patterns became increasingly localized, exhibiting a peak in the IPR near average power $P = 3.4 \text{ mW}$ (Fig. 3C). Most of the optical power in Fig. 3C was contained at the site where the light was initially launched (indicated by the red arrow). When the power was further increased, the output showed a marked delocalization (Fig. 3D), as would be expected from the numerical result presented in Fig. 2H.

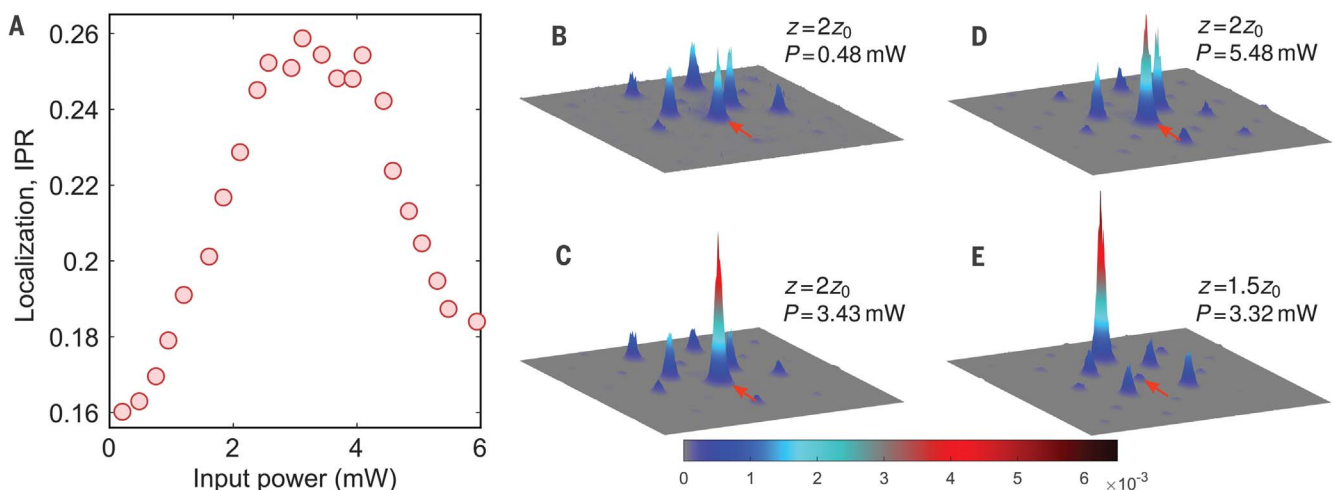


Fig. 3. Experimental observation of topological bandgap solitons. (A) IPR as a function of average input power measured at $z = 2z_0$. The presence of the peak corresponds to the existence of the topological bandgap solitons. For these measurements, \mathcal{P} at the input of the lattice was found to be 0.076 mm^{-1} per unit average input power in mW. (B to D) Corresponding output intensity distributions for three different input powers. The red arrow in each

image indicates the site where the light was launched at the input. (E) Most localized output intensity distribution measured at $z = 1.5z_0$. Note that the brightest site in this case is located directly across a diagonal from the site where the light was launched, corresponding to the cyclotron-like motion of the solitons. The field of view is smaller than the actual lattice size. Each experimentally observed intensity pattern is normalized.

In a second set of experiments, we used a similar photonic lattice with maximum propagation distance $z_{\max} = 1.5z_0$ instead of $2z_0$ (movie S5). The IPR in this case exhibited a similar behavior (i.e., delocalization to localization to delocalization) to Fig. 3A. The most localized output intensity pattern was observed near $P = 3.3$ mW (see Fig. 3E). In contrast to Fig. 3C, the brightest site in Fig. 3E was not located where the light was initially launched but was in another waveguide that was directly across a diagonal from the injected site; this constitutes direct evidence of the cyclotron-like motion of the solitons. Comparing Fig. 2H and Fig. 3A, the peak of IPR is experimentally observed at a higher power ($P = 0.26$ mm⁻¹) and the contrast of the peak is lower than what is expected from theory, as a result of linear loss. Additionally, the front and rear tails (in time) of the pulses behaved linearly, producing a small background and causing a lower contrast in IPR. That said, the observed peak in IPR is a clear signature of the topological bandgap solitons.

Our ability to control the “flatness” of the bulk band by tailoring the coupling parameter Λ is key to the observation of these solitons. In solid-state systems, flat bands play an important role in enhancing the relative strength of interactions [a recent example is twisted bilayer graphene (35)]. This is also true in our case but in a different way: The width of the linear band sets the power threshold of solitons in two dimensions. In the extreme case of a perfectly flat band, all linear Bloch states are degenerate, and thus localized eigenstates can be constructed as a superposition of these states, implying that stationary states exist even in the linear domain. Thus, in this limit, solitons have zero power threshold. Operating near the flat band with an appreciable bandgap allows us to probe the solitons at an experimentally accessible power value.

The observation of topological solitons opens a new avenue in the investigation of topological nonlinear optics, complementing other platforms such as Rydberg polaritons (36) and

nonlinear circuits (37). Furthermore, nonlinearity can act as a means to modify (38, 39) and probe (40, 41) topological photonic structures. There are many open questions; for example, the degree of robustness and the stability properties of chiral edge states in the presence of nonlinearity are unknown. It will also be of central importance to define new invariants that characterize the observable behavior of nonlinear topological systems. The interplay of nonlinearity and disorder in topological systems will be necessary if nonlinear topological devices are to be of technological use. We expect that these issues, among others, will dictate how topological states can be incorporated in useful devices based on wave mechanics, whether in the photonic, acoustic/phononic, optomechanical, atomic, polaritonic, or other domains.

REFERENCES AND NOTES

- K. von Klitzing, G. Dorda, M. Pepper, *Phys. Rev. Lett.* **45**, 494–497 (1980).
- D. J. Thouless, M. Kohmoto, M. P. Nightingale, M. den Nijs, *Phys. Rev. Lett.* **49**, 405–408 (1982).
- S. Raghu, F. D. M. Haldane, *Phys. Rev. A* **78**, 033834 (2008).
- Z. Wang, Y. Chong, J. D. Joannopoulos, M. Soljačić, *Nature* **461**, 772–775 (2009).
- M. C. Rechtsman *et al.*, *Nature* **496**, 196–200 (2013).
- M. Hafezi, S. Mittal, J. Fan, A. Migdall, J. M. Taylor, *Nat. Photonics* **7**, 1001–1005 (2013).
- M. Atala *et al.*, *Nat. Phys.* **9**, 795–800 (2013).
- G. Jotzu *et al.*, *Nature* **515**, 237–240 (2014).
- M. Aidelsburger *et al.*, *Nat. Phys.* **11**, 162–166 (2015).
- L. M. Nash *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 14495–14500 (2015).
- R. Süsstrunk, S. D. Huber, *Science* **349**, 47–50 (2015).
- J. Ningyuan, C. Owens, A. Sommer, D. Schuster, J. Simon, *Phys. Rev. X* **5**, 021031 (2015).
- T. Karzig, C.-E. Bardyn, N. H. Lindner, G. Refael, *Phys. Rev. X* **5**, 031001 (2015).
- A. V. Nalitov, D. D. Solnyshkov, G. Malpuech, *Phys. Rev. Lett.* **114**, 116401 (2015).
- L. Lu, J. D. Joannopoulos, M. Soljačić, *Nat. Photonics* **8**, 821–829 (2014).
- T. Ozawa *et al.*, *Rev. Mod. Phys.* **91**, 015006 (2019).
- A. Barthelemy, S. Maneuf, C. Froehly, *Opt. Commun.* **55**, 201–206 (1985).
- D. N. Christodoulides, R. I. Joseph, *Opt. Lett.* **13**, 794–796 (1988).
- M. Segev, B. Crosignani, A. Yariv, B. Fischer, *Phys. Rev. Lett.* **68**, 923–926 (1992).
- H. S. Eisenberg, Y. Silberberg, R. Morandotti, A. R. Boyd, J. S. Aitchison, *Phys. Rev. Lett.* **81**, 3383–3386 (1998).
- G. I. Stegeman, M. Segev, *Science* **286**, 1518–1523 (1999).
- J. W. Fleischer, M. Segev, N. K. Efremidis, D. N. Christodoulides, *Nature* **422**, 147–150 (2003).
- M. S. Rudner, N. H. Lindner, E. Berg, M. Levin, *Phys. Rev. X* **3**, 031005 (2013).
- S. Mukherjee *et al.*, *Nat. Commun.* **8**, 13918 (2017).
- L. J. Maczewsky, J. M. Zeuner, S. Nolte, A. Szameit, *Nat. Commun.* **8**, 13756 (2017).
- Y. Lumer, Y. Plotnik, M. C. Rechtsman, M. Segev, *Phys. Rev. Lett.* **111**, 243905 (2013).
- M. J. Ablowitz, C. W. Curtis, Y.-P. Ma, *Phys. Rev. A* **90**, 023813 (2014).
- D. Leykam, Y. D. Chong, *Phys. Rev. Lett.* **117**, 143901 (2016).
- J. L. Marzuola, M. Rechtsman, B. Osting, M. Bandres, arXiv 1904.10312 [cond-mat.mes-hall] (21 April 2019).
- M. Büttiker, *Phys. Rev. B* **38**, 9375–9389 (1988).
- See supplementary materials.
- S. Mukherjee, H. K. Chandrasekharan, P. Öhberg, N. Goldman, R. R. Thomson, *Nat. Commun.* **9**, 4209 (2018).
- T. Kitagawa, E. Berg, M. Rudner, E. Demler, *Phys. Rev. B* **82**, 235114 (2010).
- A. Szameit *et al.*, *Opt. Express* **14**, 6055–6062 (2006).
- Y. Cao *et al.*, *Nature* **556**, 43–50 (2018).
- L. W. Clark, N. Schine, C. Baum, N. Jia, J. Simon, arXiv 1907.05872 [cond-mat.quant-gas] (12 July 2019).
- Y. Hadad, A. B. Khanikaev, A. Alu, *Phys. Rev. B* **93**, 155112 (2016).
- D. Leykam, S. Mittal, M. Hafezi, Y. D. Chong, *Phys. Rev. Lett.* **121**, 023901 (2018).
- L. He *et al.*, *Nat. Commun.* **10**, 4194 (2019).
- S. Kruk *et al.*, *Nat. Nanotechnol.* **14**, 126–130 (2019).
- D. Smirnova *et al.*, *Phys. Rev. Lett.* **123**, 103901 (2019).

ACKNOWLEDGMENTS

We thank H. K. Chandrasekharan, J. Guglielmon, D. Leykam, and J. Noh for useful discussions and N. C. Giebink for use of a supercontinuum laser source for directional coupler characterization. **Funding:** Supported by Office of Naval Research award N00014-18-1-2595 (S.M. and M.C.R.) and by Packard Foundation fellowship 2017-66821 and Kaufman Foundation award KA2017-91788 (M.C.R.). **Author contributions:** S.M. designed and built the waveguide fabrication system as well as characterization setups and carried out all experiments; S.M. and M.C.R. conceived the idea, designed the experiment, analyzed data, and wrote the manuscript; S.M. performed theoretical analysis with input from M.C.R.; M.C.R. supervised the project. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All associated data and materials are available in the manuscript and supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6493/856/suppl/DC1
Supplementary Text
Figs. S1 to S9
Movies S1 to S5
References (42–46)

12 January 2020; accepted 14 April 2020
10.1126/science.aba8725

CORONAVIRUS

Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period

Stephen M. Kissler^{1*}, Christine Tedijanto^{2*}, Edward Goldstein², Yonatan H. Grad^{1†‡}, Marc Lipsitch^{2†‡}

It is urgent to understand the future of severe acute respiratory syndrome–coronavirus 2 (SARS-CoV-2) transmission. We used estimates of seasonality, immunity, and cross-immunity for human coronavirus OC43 (HCoV-OC43) and HCoV-HKU1 using time-series data from the United States to inform a model of SARS-CoV-2 transmission. We projected that recurrent wintertime outbreaks of SARS-CoV-2 will probably occur after the initial, most severe pandemic wave. Absent other interventions, a key metric for the success of social distancing is whether critical care capacities are exceeded. To avoid this, prolonged or intermittent social distancing may be necessary into 2022. Additional interventions, including expanded critical care capacity and an effective therapeutic, would improve the success of intermittent distancing and hasten the acquisition of herd immunity. Longitudinal serological studies are urgently needed to determine the extent and duration of immunity to SARS-CoV-2. Even in the event of apparent elimination, SARS-CoV-2 surveillance should be maintained because a resurgence in contagion could be possible as late as 2024.

The ongoing severe acute respiratory syndrome–coronavirus 2 (SARS-CoV-2) pandemic has caused nearly 500,000 detected cases of coronavirus disease 2019 (COVID-19) illness and claimed >20,000 lives worldwide as of 26 March 2020 (1). Experience from China, Italy, and the United States demonstrates that COVID-19 can overwhelm even the healthcare capacities of well-resourced nations (2–4). With no pharmaceutical treatments available, interventions have focused on contact tracing, quarantine, and social distancing. The required intensity, duration, and urgency of these responses will depend both on how the initial pandemic wave unfolds and on the subsequent transmission dynamics of SARS-CoV-2. During this initial pandemic wave, many countries have adopted social distancing measures and some, like China, are gradually lifting them after achieving adequate control of transmission. However, to mitigate the possibility of resurgences of infection, prolonged or intermittent periods of social distancing may be required. After the initial pandemic wave, SARS-CoV-2 might follow its closest genetic relative, SARS-CoV-1, and be eradicated by intensive public health measures after causing a brief but intense pandemic (5). Increasingly, public health authorities consider this scenario unlikely (6). Alternatively, the transmission of SARS-CoV-2 could resemble that of pandemic influenza by circulating seasonally after causing an initial global wave

of infection (7). Such a scenario could reflect the previous emergence of known human coronaviruses (HCoVs) from zoonotic origins, e.g., HCoV-OC43 (8). Distinguishing between these scenarios is key for formulating an effective, sustained public health response to SARS-CoV-2.

The pandemic and postpandemic transmission dynamics of SARS-CoV-2 will depend on factors including the degree of seasonal variation in transmission, the duration of immunity, and the degree of cross-immunity between SARS-CoV-2 and other coronaviruses, as well as the intensity and timing of control measures. SARS-CoV-2 belongs to the *Betacoronavirus* genus, which includes the SARS-CoV-1 coronavirus, the Middle East respiratory syndrome (MERS) coronavirus, and two other HCoVs, HCoV-OC43 and HCoV-HKU1. The SARS-CoV-1 and MERS coronaviruses cause severe illness with approximate case fatality rates of 9 and 36%, respectively, but the transmission of both has remained limited (9). HCoV-OC43 and HCoV-HKU1 infections may be asymptomatic or associated with mild to moderate upper respiratory tract illness; these HCoVs are considered the second most common cause of the common cold (9). HCoV-OC43 and HCoV-HKU1 cause annual wintertime outbreaks of respiratory illness in temperate regions (10, 11), suggesting that wintertime climate and host behaviors may facilitate transmission, as is true for influenza (12–14). Immunity to HCoV-OC43 and HCoV-HKU1 appears to wane appreciably within 1 year (15), whereas SARS-CoV-1 infection can induce longer-lasting immunity (16). The betacoronaviruses can induce immune responses against one another: SARS-CoV-1 infection can generate neutralizing antibodies against HCoV-OC43 (16) and HCoV-OC43 infection can generate cross-reactive antibodies

against SARS-CoV-1 (17). Although investigations into the spectrum of illness caused by SARS-CoV-2 are ongoing, recent evidence indicates that most patients experience mild to moderate illness with more limited occurrence of severe lower respiratory infection (18). Current COVID-19 case fatality rates are estimated to lie between 0.6 and 3.5% (19, 20), suggesting lower severity than SARS-CoV-1 and MERS but higher severity than HCoV-OC43 and HCoV-HKU1. The high infectiousness near the start of often mild symptoms makes SARS-CoV-2 considerably harder to control with case-based interventions such as intensive testing, isolation, and tracing compared with the SARS-CoV-1 and MERS coronaviruses (21).

Intensive testing and case-based interventions have so far formed the centerpiece of control efforts in some places, including Singapore and Hong Kong (22). Many other countries are adopting measures such as social distancing, closing schools and workplaces, and limiting the sizes of gatherings. The goal of these strategies is to reduce the peak intensity of the pandemic (i.e., “flatten the curve”) (22), reducing the risk of overwhelming health systems and buying time to develop treatments and vaccines. For social distancing to have reversed the pandemic in China, the effective reproduction number (R_e ; defined as the average number of secondary infections caused by a single infected individual in the population after there is some immunity or interventions have been put in place) must have declined by at least 50 to 60%, assuming a baseline basic reproduction number (R_0 ; defined as the average number of secondary infections caused by a single infected individual in a completely susceptible population) between 2 and 2.5 (22). Through intensive control measures, Shenzhen was able to reduce the R_e by an estimated 85% (23). However, it is unclear how well these declines in R_0 might generalize to other settings: recent data from Seattle suggest that the R_0 has only declined to about 1.4, or by about 30 to 45%, assuming a baseline R_0 between 2 and 2.5 (24). Furthermore, social distancing measures may need to last for months to effectively control transmission and mitigate the possibility of resurgence (25).

A key metric for the success of social distancing interventions is whether critical care capacities are exceeded. Modeling studies (26) and experience from the Wuhan outbreak (2) indicate that critical care capacities even in high-income countries can be exceeded many times over if distancing measures are not implemented quickly or strongly enough. To alleviate these problems, approaches to increasing critical care capacity have included rapid construction or repurposing of hospital facilities and consideration of increased manufacturing and distribution of ventilators

¹Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, MA, USA.

²Department of Epidemiology, Harvard T.H. Chan School of Public Health, Boston, MA, USA.

*These authors contributed equally to this work and are cosenior authors. †These authors contributed equally to this work.

‡Corresponding author. Email: mlipsitch@hsph.harvard.edu (M.L.); ygrad@hsph.harvard.edu (Y.H.G.)

(27–30). Treatments that reduce the proportion of infections that lead to severe illness could have a similar effect of reducing burden on healthcare systems.

Here, we identify viral, environmental, and immunologic factors that in combination will determine the dynamics of SARS-CoV-2. We integrate our findings in a mathematical model to project potential scenarios for SARS-CoV-2 transmission through the pandemic and post-pandemic periods and identify key data still needed to determine which scenarios are likely to play out. Then, using the model, we assess the duration and intensity of social distancing measures that might be needed to maintain control of SARS-CoV-2 in the coming months under both existing and expanded critical care capacities.

Transmission dynamics of HCoV-OC43 and HCoV-HKU1

We used data from the United States to model betacoronavirus transmission in temperate regions and to project the possible dynamics of SARS-CoV-2 infection through the year 2025. We first assessed the role of seasonal variation, duration of immunity, and cross-immunity on the transmissibility of HCoV-OC43 and HCoV-HKU1 in the United States. We used the weekly percentage of positive laboratory tests for HCoV-OC43 and HCoV-HKU1 (31) multiplied by the weekly population-weighted proportion of physician visits for influenza-like illness (ILI) (32, 33) to approximate historical betacorona-

virus incidence in the United States to within a scaling constant. This proxy is proportional to incidence under a set of assumptions described in the supplementary materials and methods. To quantify variation in transmission strength over time, we estimated the weekly R_e (34, 35). The R_e s for each of the betacoronaviruses displayed a seasonal pattern, with annual peaks in the R_e slightly preceding those of the incidence curves (fig. S1). We limited our analysis to “in-season” estimates that were based on adequate samples, defined as week 40 through week 20 of the following year, roughly October to May. For both HCoV-OC43 and HCoV-HKU1, the R_e typically reached its peak between October and November and its trough between February and May. Over the five seasons included in our data (2014 to 2019), the median peak R_e was 1.85 (range: 1.61 to 2.21) for HCoV-HKU1 and 1.56 (range: 1.54 to 1.80) for HCoV-OC43 after removing outliers (five for HCoV-HKU1, zero for HCoV-OC43). Results were similar using various choices of incidence proxy and serial interval distributions (figs. S1 to S3).

To quantify the relative contribution of immunity versus seasonal forcing on the transmission dynamics of the betacoronaviruses, we adapted a regression model (36) that expressed the R_e for each strain (HKU1 and OC43) as the product of a baseline transmissibility constant (related to the R_0) and the proportion of the population susceptible (hereafter referred to as “susceptibles”) at the start of each season, the depletion of susceptibles because of in-

fection with the same strain, the depletion of susceptibles because of infection with the other strain, and a spline to capture further unexplained seasonal variation in transmission strength (seasonal forcing). These covariates were able to explain most of the observed variability in the R_e s (adjusted R^2 : 74.3%). The estimated multiplicative effects of each of these covariates on the weekly R_e are depicted in Fig. 1. As expected, depletion of susceptibles for each betacoronavirus strain was negatively correlated with transmissibility of that strain. Depletion of susceptibles for each strain was also negatively correlated with the R_e of the other strain, providing evidence of cross-immunity. Per incidence proxy unit, the effect of the cross-immunizing strain was always less than the effect of the strain itself (table S1), but the overall impact of cross-immunity on the R_e could still be substantial if the cross-immunizing strain had a large outbreak (e.g., HCoV-OC43 in 2014–2015 and 2016–2017). The ratio of cross-immunization to self-immunization effects was larger for HCoV-HKU1 than for HCoV-OC43, suggesting that HCoV-OC43 confers stronger cross-immunity. Seasonal forcing appears to drive the rise in transmissibility at the start of the season (late October through early December), whereas depletion of susceptibles plays a comparatively larger role in the decline in transmissibility toward the end of the season. The strain-season coefficients were fairly consistent across seasons for each strain and lacked a

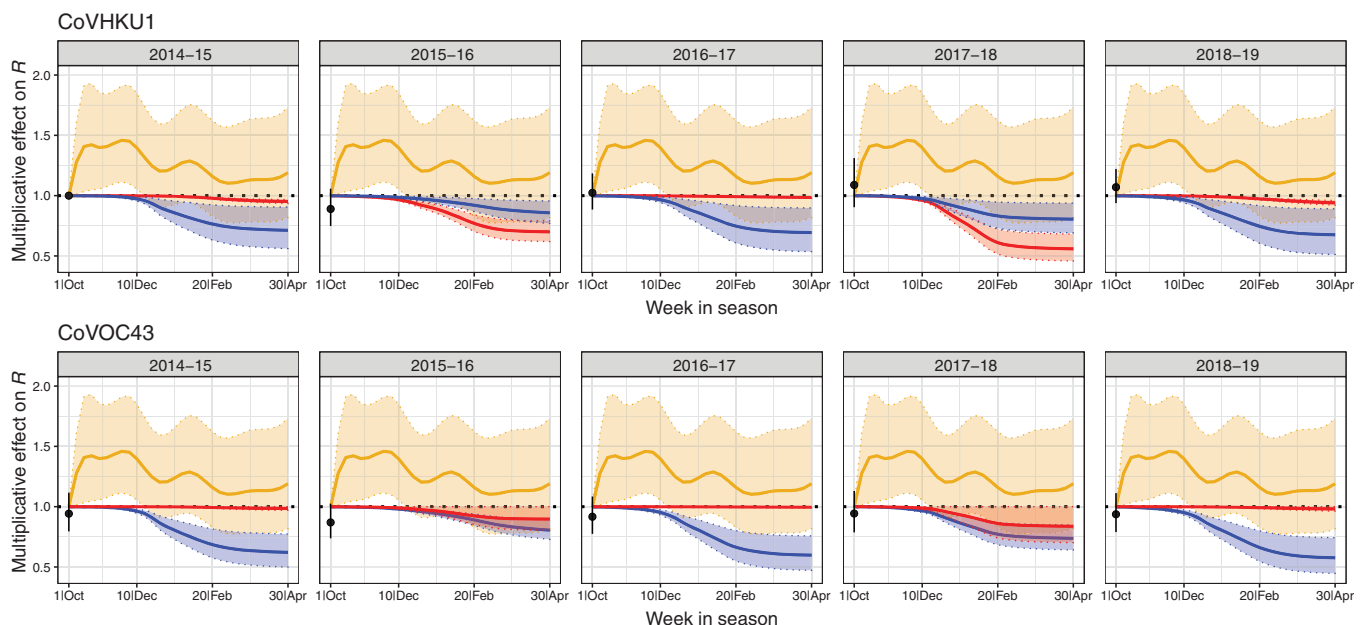


Fig. 1. Effects of depletion of susceptibles and seasonality on R_e by strain and season. Shown are the estimated multiplicative effects of HCoV-HKU1 incidence (red), HCoV-OC43 incidence (blue), and seasonal forcing (gold) on weekly R_e s of HCoV-HKU1 (top) and HCoV-OC43 (bottom), with 95% confidence intervals. The black dot (with 95% confidence interval) plotted at the start of each season is the estimated coefficient for that strain and season compared with the 2014–2015 HCoV-HKU1 season. The seasonal forcing spline is set to 1 at the first week of the season (no intercept). On the x-axis, the first “week in season” corresponds to epidemiological week 40.

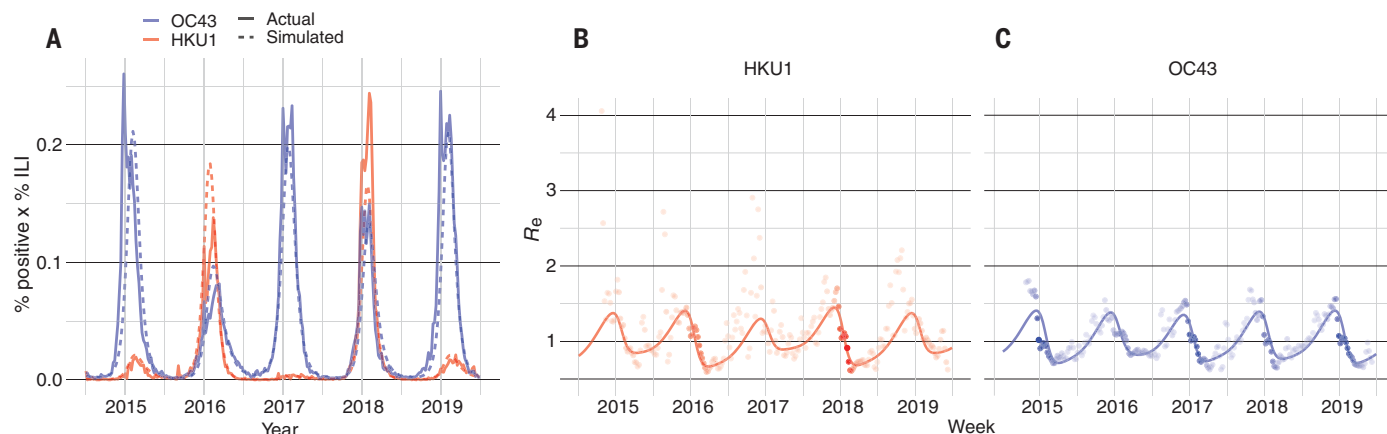


Fig. 2. Transmission model fits for HCoV-OC43 and HCoV-HKU1. (A) Weekly percent positive laboratory tests multiplied by percent ILI for HCoV-OC43 (blue) and HCoV-HKU1 (red) in the United States between 5 July 2014 and 29 June 2019 (solid lines) with simulated output from the best-fit SEIRS transmission model (dashed lines). (B and C) Weekly R_e values estimated using the Wallinga–Teunis method (points) and simulated R_e from the best-fit SEIRS

transmission model (line) for HCoV-OC43 and HCoV-HKU1. The opacity of each point is determined by the relative percent ILI multiplied by percent positive laboratory tests in that week relative to the maximum percent ILI multiplied by percent positive laboratory tests for that strain across the study period, which reflects uncertainty in the R_e estimate; estimates are more certain (darker points) in weeks with higher incidence.

clear correlation with incidence in prior seasons, consistent with experimental results showing substantial waning of immunity within 1 year (15).

We integrated these findings into a two-strain ordinary differential equation susceptible-exposed-infectious-recovered-susceptible (SEIRS) compartmental model to describe the transmission dynamics of HCoV-OC43 and HCoV-HKU1 (fig. S4). The model provided a good fit to both the weekly incidence proxies for HCoV-OC43 and HCoV-HKU1 and to the estimated weekly R_e s (Fig. 2). According to the best-fit model parameters, the R_0 for HCoV-OC43 and HCoV-HKU1 varies between 1.7 in the summer and 2.2 in the winter and peaks in the second week of January, consistent with the seasonal spline estimated from the data. Also in agreement with the findings of the regression model, the duration of immunity for both strains in the best-fit SEIRS model is ~45 weeks, and each strain induces cross-immunity against the other, although the cross-immunity that HCoV-OC43 infection induces against HCoV-HKU1 is stronger than the reverse.

Simulating the transmission of SARS-CoV-2

Next, we incorporated a third betacoronavirus into the dynamic transmission model to represent SARS-CoV-2. We assumed a latent period of 4.6 days (26, 37–39) and an infectious period of 5 days, informed by the best-fit values for the other betacoronaviruses (table S8). We allowed the cross-immunities, duration of immunity, maximum R_0 , and degree of seasonal variation in R_0 to vary. We assumed an establishment time of sustained transmission on 11 March 2020, when the World Health Orga-

nization declared the SARS-CoV-2 outbreak a pandemic (40), and we varied the establishment time in a sensitivity analysis (fig. S7). For a representative set of parameter values, we determined annual SARS-CoV-2 infections (tables S2 to S4 and fig. S7) and the peak annual SARS-CoV-2 prevalence (tables S5 to S7 and fig. S7) through 2025. We summarized the postpandemic SARS-CoV-2 dynamics into the categories of annual outbreaks, biennial outbreaks, sporadic outbreaks, or virtual elimination (tables S2 to S7). Overall, shorter durations of immunity and smaller degrees of cross-immunity from the other betacoronaviruses were associated with greater total incidence of infection by SARS-CoV-2, and autumn establishments and smaller seasonal fluctuations in transmissibility were associated with larger pandemic peak sizes. Model simulations demonstrated the following key points.

SARS-CoV-2 can proliferate at any time of year

In all modeled scenarios, SARS-CoV-2 was capable of producing a substantial outbreak regardless of establishment time. Spring/summer establishments favored outbreaks with lower peaks, whereas autumn/winter establishments led to more acute outbreaks (tables S5 to S7 and fig. S7). The 5-year cumulative incidence proxies were comparable for all establishment times (tables S5 to S7).

If immunity to SARS-CoV-2 is not permanent, it will likely enter into regular circulation

Much like pandemic influenza, many scenarios lead to SARS-CoV-2 entering into long-term circulation alongside the other human betacoronaviruses (e.g., Fig. 3, A and B), possibly in annual, biennial, or sporadic patterns, over

the next 5 years (tables S2 to S4). Short-term immunity (~40 weeks, similar to HCoV-OC43 and HCoV-HKU1) favors the establishment of annual SARS-CoV-2 outbreaks, whereas longer-term immunity (2 years) favors biennial outbreaks.

High seasonal variation in transmission leads to smaller peak incidence during the initial pandemic wave but larger recurrent wintertime outbreaks

The amount of seasonal variation in SARS-CoV-2 transmission could differ between geographic locations, as is the case for influenza (12). The R_0 for influenza in New York declines in the summer by ~40%, whereas in Florida the decline is closer to 20%, which aligns with the estimated decline in R_0 for HCoV-OC43 and HCoV-HKU1 (table S8). A 40% summertime decline in R_0 would reduce the unmitigated peak incidence of the initial SARS-CoV-2 pandemic wave. However, stronger seasonal forcing leads to a greater accumulation of susceptible individuals during periods of low transmission in the summer, leading to recurrent outbreaks with higher peaks in the postpandemic period (Fig. 3C).

If immunity to SARS-CoV-2 is permanent, the virus could disappear for 5 or more years after causing a major outbreak

Long-term immunity consistently led to effective elimination of SARS-CoV-2 and a lower overall incidence of infection. If SARS-CoV-2 induces cross-immunity against HCoV-OC43 and HCoV-HKU1, then the incidence of all betacoronaviruses could decline and even virtually disappear (Fig. 3D). The virtual elimination of HCoV-OC43 and HCoV-HKU1 would

be possible if SARS-CoV-2 induced 70% cross-immunity against them, which is the same estimated level of cross-immunity that HCoV-OC43 induces against HCoV-HKU1.

Low levels of cross-immunity from the other betacoronaviruses against SARS-CoV-2 could make SARS-CoV-2 appear to die out, only to resurge after a few years

Even if SARS-CoV-2 immunity only lasts for 2 years, mild (30%) cross-immunity from HCoV-OC43 and HCoV-HKU1 could effectively eliminate the transmission of SARS-CoV-2 for up to 3 years before a resurgence in 2024, as long as SARS-CoV-2 does not fully die out (Fig. 3E).

To illustrate these scenarios (Fig. 3), we used a maximum wintertime R_0 of 2.2, informed by the estimated R_0 for HCoV-OC43 and HCoV-HKU1 (table S8). This is a low but plausible estimate of the R_0 for SARS-CoV-2 (41). Increasing the wintertime R_0 to 2.6 leads to more intense outbreaks but the qualitative range of scenarios remains similar (fig. S8).

Assessing intervention scenarios during the initial pandemic wave

Regardless of the postpandemic transmission dynamics of SARS-CoV-2, urgent measures are required to address the ongoing pandemic. Pharmaceutical treatments and vaccines may require months to years to develop and test, leaving nonpharmaceutical interventions as the only immediate means of curbing SARS-CoV-2 transmission. Social distancing measures have been adopted in many countries with widespread SARS-CoV-2 transmission. The necessary duration and intensity of these measures has yet to be characterized. To address this, we adapted the SEIRS transmission model (fig. S9) to capture moderate, mild, or asymptomatic infections (95.6% of infections), infections that lead to hospitalization but not critical care (3.08%), and infections that require critical care (1.32%) (26). We assumed the worst-case scenario of no cross-immunity from HCoV-OC43 and HCoV-HKU1 against SARS-CoV-2, which makes the SARS-CoV-2 model unaffected by the transmission dynamics of

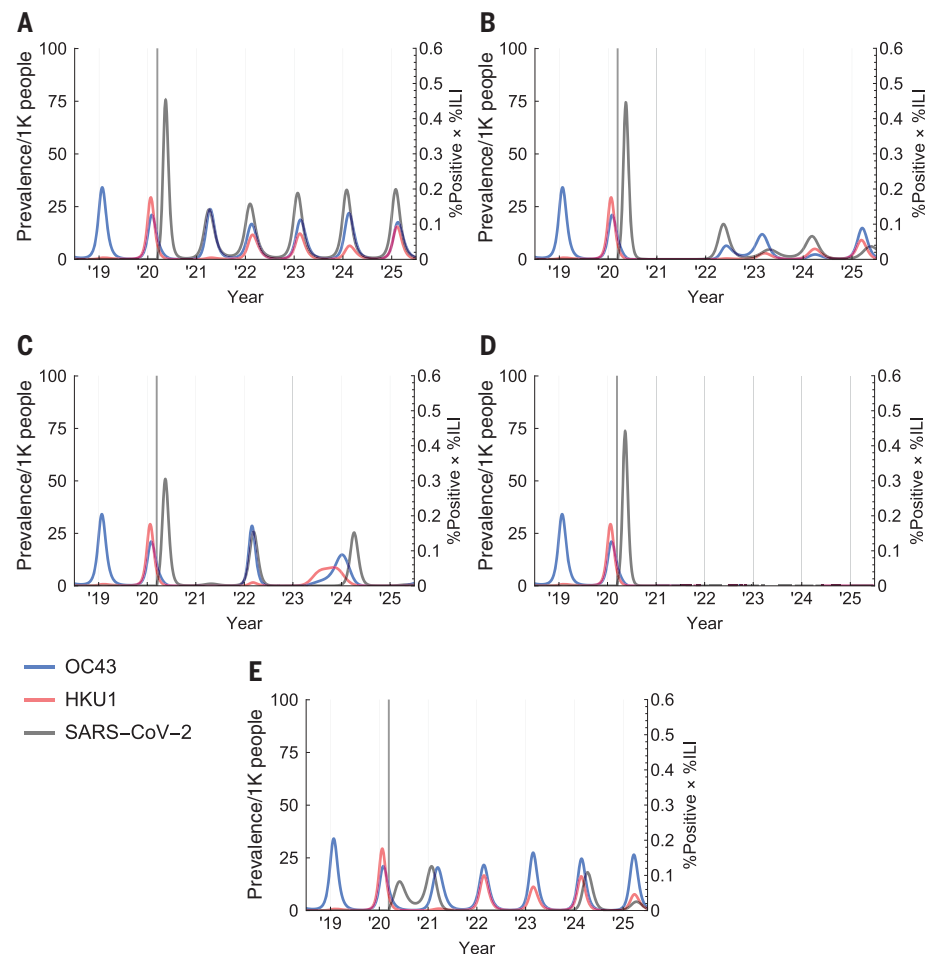
those viruses. Informed by the transmission model fits, we assumed a latent period of 4.6 days and an infectious period of 5 days, in agreement with estimates from other studies (26). The mean duration of noncritical hospital stay was 8 days for those not requiring critical care and 6 days for those requiring critical care, and the mean duration of critical care was 10 days (26). We varied the peak (wintertime) R_0 between 2.2 and 2.6 and allowed the summertime R_0 to vary between 60% (i.e., relatively strong seasonality) and 100% (i.e., no seasonality) of the wintertime R_0 , guided by the inferred seasonal forcing for HCoV-OC43 and HCoV-HKU1 (table S8).

We used the open critical care capacity of the United States, 0.89 free beds per 10,000 adults, as a benchmark for critical care demand (2). We simulated pandemic trajectories that were based on a pandemic establishment time of 11 March 2020. We simulated social distancing by reducing R_0 by a fixed proportion, which ranged between 0 and 60%. We assessed “one-time” social distancing interventions, for which

Fig. 3. Invasion scenarios for SARS-CoV-2 in temperate regions.

These plots depict the prevalence of SARS-CoV-2 (black, cases per 1000 people), HCoV-OC43 (blue, percent positive multiplied by percent ILI), and HCoV-HKU1 (red, percent positive multiplied by percent ILI) for a representative set of possible pandemic and postpandemic scenarios. The scenarios were obtained by varying the cross-immunity between SARS-CoV-2 and HCoVs OC43/HKU1 (χ_{3X}) and vice versa (χ_{X3}), the duration of SARS-CoV-2 immunity ($1/\sigma_3$), and the seasonal variation in R_0 (f), assuming an pandemic establishment time of 11 March 2020 (depicted as a vertical gray bar). Parameter values used to generate each plot are listed below; all other parameters were held at the values listed in table S8.

(A) A short duration ($1/\sigma_3 = 40$ weeks) of SARS-CoV-2 immunity could yield annual SARS-CoV-2 outbreaks. (B) Longer-term SARS-CoV-2 immunity ($1/\sigma_3 = 104$ weeks) could yield biennial outbreaks, possibly with smaller outbreaks in the intervening years. (C) Higher seasonal variation in transmission ($f = 0.4$) would reduce the peak size of the invasion wave but could lead to more severe wintertime outbreaks thereafter [compare with (B)]. (D) Long-term immunity ($1/\sigma_3 = \text{infinity}$) to SARS-CoV-2 could lead to elimination of the virus. (E) However, a resurgence of SARS-CoV-2 could occur as late as 2024 after a period of apparent elimination if the duration of immunity is intermediate ($1/\sigma_3 = 104$ weeks) and if HCoV-OC43 and HCoV-HKU1 impart intermediate cross-immunity against SARS-CoV-2 ($\chi_{3X} = 0.3$). (A) $\chi_{3X} = 0.3$, $\chi_{X3} = 0$, $1/\sigma_3 = 40$ weeks, $f = 0.2$. (B) $\chi_{3X} = 0.7$, $\chi_{X3} = 0$, $1/\sigma_3 = 104$ weeks, $f = 0.2$. (C) $\chi_{3X} = 0.7$, $\chi_{X3} = 0$, $1/\sigma_3 = 104$ weeks, $f = 0.4$. (D) $\chi_{3X} = 0.7$, $\chi_{X3} = 0$, $1/\sigma_3 = \text{infinity}$, $f = 0.2$. (E) $\chi_{3X} = 0.3$, $\chi_{X3} = 0.3$, $1/\sigma_3 = 104$ weeks, $f = 0.4$.



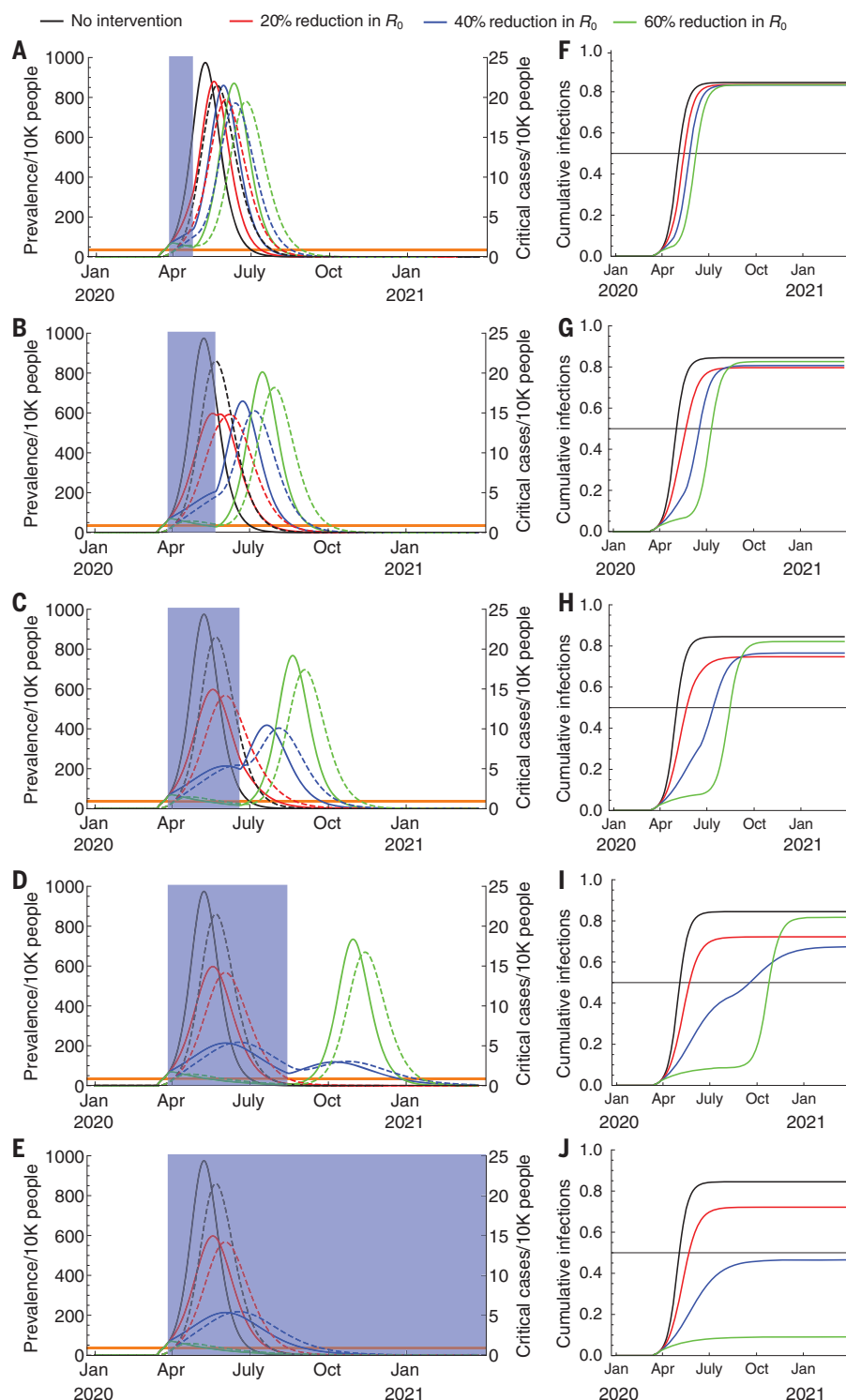


Fig. 4. One-time social distancing scenarios in the absence of seasonality. (A to E) Simulated prevalence of COVID-19 infections (solid) and critical COVID-19 cases (dashed) after establishment on 11 March 2020 with a period of social distancing (shaded blue region) instated 2 weeks later, with the duration of social distancing lasting (A) 4 weeks, (B) 8 weeks, (C) 12 weeks, (D) 20 weeks, and (E) indefinitely. There is no seasonal forcing; R_0 was held constant at 2.2 (see fig. S12 for $R_0 = 2.6$). The effectiveness of social distancing varied from none to a 60% reduction in R_0 . Cumulative infection sizes are depicted beside each prevalence plot (F to J) with the herd immunity threshold (horizontal black bar). Of the temporary distancing scenarios, long-term (20-week), moderately effective (20 to 40%) social distancing yields the smallest overall peak and total outbreak size.

R_0 was reduced by up to 60% for a fixed duration of time (up to 20 weeks) or indefinitely starting 2 weeks after pandemic establishment. We also assessed intermittent social distancing measures, for which social distancing was turned “on” when the prevalence of infection rose above a threshold and “off” when it fell below a second, lower threshold, with the goal of keeping the number of critical care patients below 0.89 per 10,000 adults. An “on” threshold of 35 cases per 10,000 people achieved this goal in both the seasonal and nonseasonal cases with wintertime $R_0 = 2.2$. We chose five cases per 10,000 adults as the “off” threshold. These thresholds were chosen to qualitatively illustrate the intermittent intervention scenario; in practice, the thresholds will need to be tuned to local epidemic dynamics and hospital capacities. We performed a sensitivity analysis around these threshold values (figs. S10 and S11) to assess how they affected the duration and frequency of the interventions. We also implemented a model with extra compartments for the latent period, infectious period, and each hospitalization period so that the waiting times in these states were gamma distributed instead of being exponentially distributed (see the supplementary materials and methods and figs. S16 and S17). Finally, we assessed the impact of doubling critical care capacity (and the associated on/off thresholds) on the frequency and overall duration of the social distancing measures.

We evaluated the impact of one-time social distancing efforts of varying effectiveness and duration on the peak and timing of the pandemic with and without seasonal forcing. When transmission was not subject to seasonal forcing, one-time social distancing measures reduced the pandemic peak size (Fig. 4 and fig. S12). Under all scenarios, there was a resurgence of infection when the simulated social distancing measures were lifted. However, longer and more stringent temporary social distancing did not always correlate with greater reductions in pandemic peak size. In the case of a 20-week period of social distancing with a 60% reduction in R_0 , for example (Fig. 4D), the resurgence peak size was nearly the same as the peak size of the uncontrolled pandemic: the social distancing was so effective that virtually no population immunity was built. The greatest reductions in peak size come from social distancing intensity and duration that divide cases approximately equally between peaks (42).

For simulations with seasonal forcing, the postintervention resurgent peak could exceed the size of the unconstrained pandemic (Fig. 5 and fig. S13), both in terms of peak prevalence and in terms of total number infected. Strong social distancing maintains a high proportion of susceptible individuals in the population, leading to an intense resurgence when R_0 rises

in the late autumn and winter. None of the one-time interventions was effective at maintaining the prevalence of critical cases below the critical care capacity.

Intermittent social distancing could prevent critical care capacity from being exceeded (Fig. 6 and fig. S14). Because of the natural history of infection, there is an ~3-week lag between the start of social distancing and the peak critical care demand. When transmission is seasonally forced, summertime social distancing can be less frequent than when R_0 remains constant at its maximal wintertime value throughout the year. The length of time between distancing measures increases as the pandemic continues because the accumulation of immunity in the population slows the resurgence of infection. Under current critical care capacities, however, the overall duration of the SARS-CoV-2 pandemic could last into 2022, requiring social distancing measures to be in place between 25% (for wintertime $R_0 = 2$ and seasonality; fig. S11A) and 75% (for wintertime $R_0 = 2.6$ and no seasonality; fig. S10C) of that time. When the latent, infectious, and hospitalization periods are gamma distributed, incidence rises more quickly, requiring a lower threshold for implementing distancing measures (25 cases per 10,000 individuals for $R_0 = 2.2$ in our model) and more frequent interventions (fig. S16).

Increasing critical care capacity allows population immunity to be accumulated more rapidly, reducing the overall duration of the pandemic and the total length of social distancing measures (Fig. 6, C and D). Although the frequency and duration of the social distancing measures were similar between the scenarios with current and expanded critical care capacity, the pandemic would conclude by July 2022 and social distancing measures could be fully relaxed by early to mid-2021, depending again on the degree of seasonal forcing of transmission (Fig. 6, C and D). Introducing a hypothetical treatment that halved the proportion of infections that required hospitalization had a similar effect as doubling critical care capacity (fig. S15).

Discussion

Here, we examined a range of likely SARS-CoV-2 transmission scenarios through 2025 and assessed nonpharmaceutical interventions that could mitigate the intensity of the current outbreak. If immunity to SARS-CoV-2 wanes in the same manner as related coronaviruses, then recurrent wintertime outbreaks are likely to occur in coming years. The total incidence of SARS-CoV-2 through 2025 will depend crucially on this duration of immunity and, to a lesser degree, on the amount of cross-immunity that exists between HCoV-OC43/HCoV-HKU1 and SARS-CoV-2. The intensity of the initial pandemic wave will depend fundamentally on

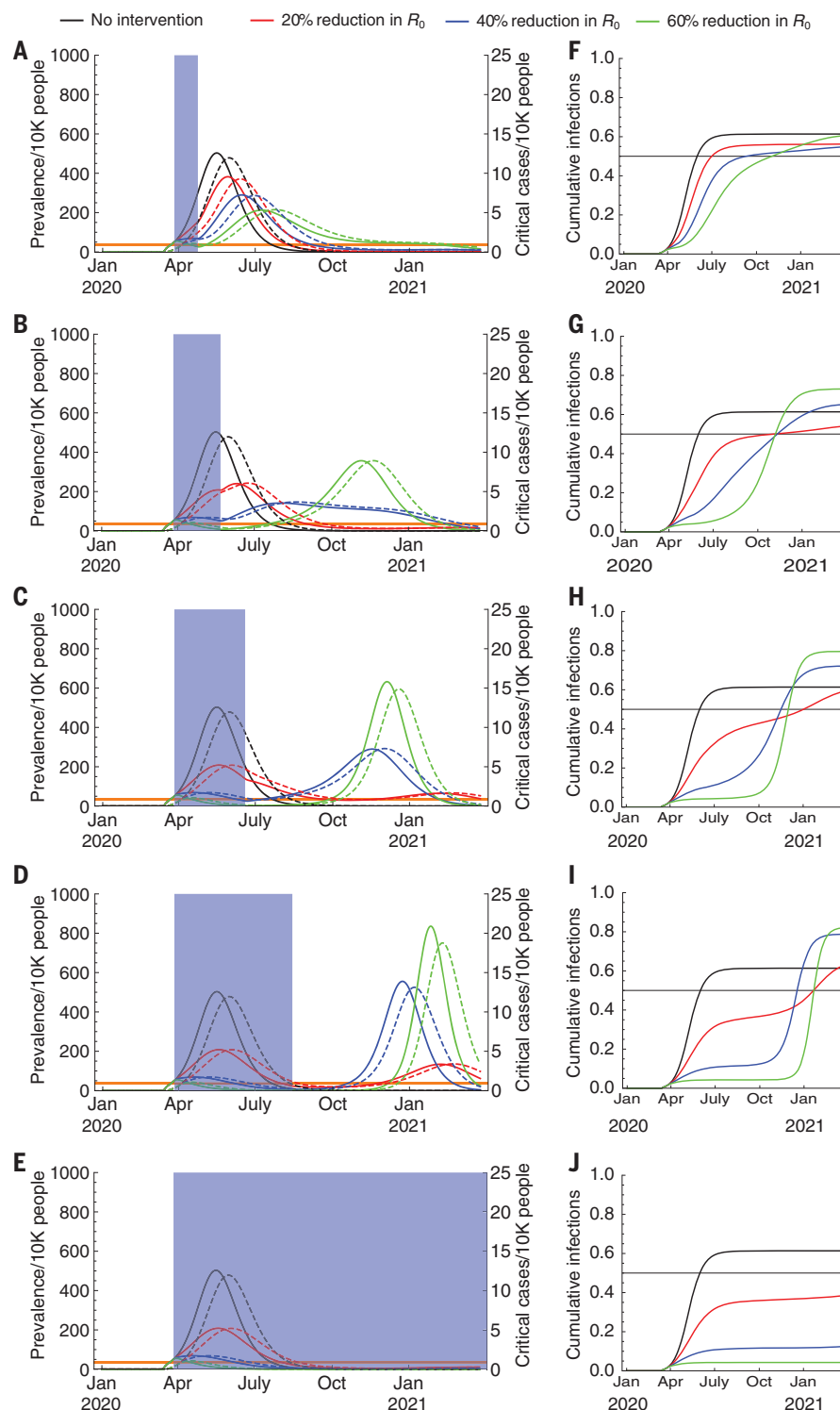


Fig. 5. One-time social distancing scenarios with seasonal transmission. (A to E) Simulated prevalence assuming strong seasonal forcing (wintertime $R_0 = 2.2$, summertime $R_0 = 1.3$, a 40% decline) of COVID-19 infections (solid) and critical COVID-19 cases (dashed) after establishment on 11 March 2020 with a period of social distancing (shaded blue region) instated 2 weeks later, with the duration of social distancing lasting (A) 4 weeks, (B) 8 weeks, (C) 12 weeks, (D) 20 weeks, and (E) indefinitely (see fig. S13 for a scenario with wintertime $R_0 = 2.6$). The effectiveness of social distancing varied from none to a 60% reduction in R_0 . Cumulative infection sizes are depicted beside each prevalence plot (F to J) with the herd immunity threshold (horizontal black bar). Preventing widespread infection during the summer can flatten and prolong the pandemic but can also lead to a high density of susceptible individuals who could become infected in an intense autumn wave.

the R_0 at the time of pandemic establishment: If establishment occurs in the autumn when the R_e is rising, which could occur in countries that maintain pandemic control by contact tracing and quarantine through the summer, or if SARS-CoV-2 is not subject to the same summertime decline in transmissibility as HCoV-OC43 and HCoV-HKU1, then a high peak prevalence of infection is likely. One-time social distancing efforts may push the SARS-CoV-2 pandemic peak into the autumn, potentially exacerbating the load on critical

care resources if there is increased wintertime transmissibility. Intermittent social distancing might maintain critical care demand within current thresholds, but widespread surveillance will be required to time the distancing measures correctly and avoid overshooting critical care capacity. New therapeutics, vaccines, or other interventions such as aggressive contact tracing and quarantine—impractical now in many places but more practical once case numbers have been reduced and testing scaled up (43)—could alleviate the need for stringent

social distancing to maintain control of the pandemic. In the absence of such interventions, surveillance and intermittent distancing (or sustained distancing if it is highly effective) may need to be maintained into 2022, which would present a substantial social and economic burden. To shorten the SARS-CoV-2 pandemic and to ensure adequate care for the critically ill, increasing critical care capacity and developing additional interventions are urgent priorities. Meanwhile, serological testing is required to understand the extent and duration of immunity to SARS-CoV-2, which will help to determine the postpandemic dynamics of the virus. Sustained, widespread surveillance will be needed both in the short term to effectively implement intermittent social distancing measures and in the long term to assess the possibility of resurgences of SARS-CoV-2 infection, which could occur as late as 2025 even after a prolonged period of apparent elimination.

Our observations are consistent with other predictions of how SARS-CoV-2 transmission might unfold and with assessments of the mitigation efforts that might be needed to curb the current outbreak. A modeling study using data from Sweden found that seasonal establishment of SARS-CoV-2 transmission is likely in the postpandemic period (11). Observational and modeling studies (2, 26) have found that early implementation of strong social distancing is essential for controlling the spread of SARS-CoV-2 and that, in the absence of the development of new therapies or preventative measures such as aggressive case finding and quarantining (21), intermittent distancing measures may be the only way to avoid overwhelming critical care capacity while building population immunity. The observation that strong, temporary social distancing can lead to especially large resurgences agrees with data from the 1918 influenza pandemic in the United States (44), in which the size of the autumn 1918 peak of infection was inversely associated with that of a subsequent winter peak after interventions were no longer in place.

Our study was subject to a variety of limitations. Only five seasons of observational data on coronaviruses were available, although the incidence patterns resemble those from 10 years of data from a hospital in Sweden (11). We assumed that the spline coefficients were constant across all seasons but seasonal forcing likely differed from year to year because of underlying drivers. To keep the transmission model from becoming unreasonably complex, we assumed that there was no difference in the seasonal forcing, per-case force of infection, latent period, or infectious period across betacoronaviruses. However, our estimates for these values lie within the range of estimates from the literature. Although disease dynamics

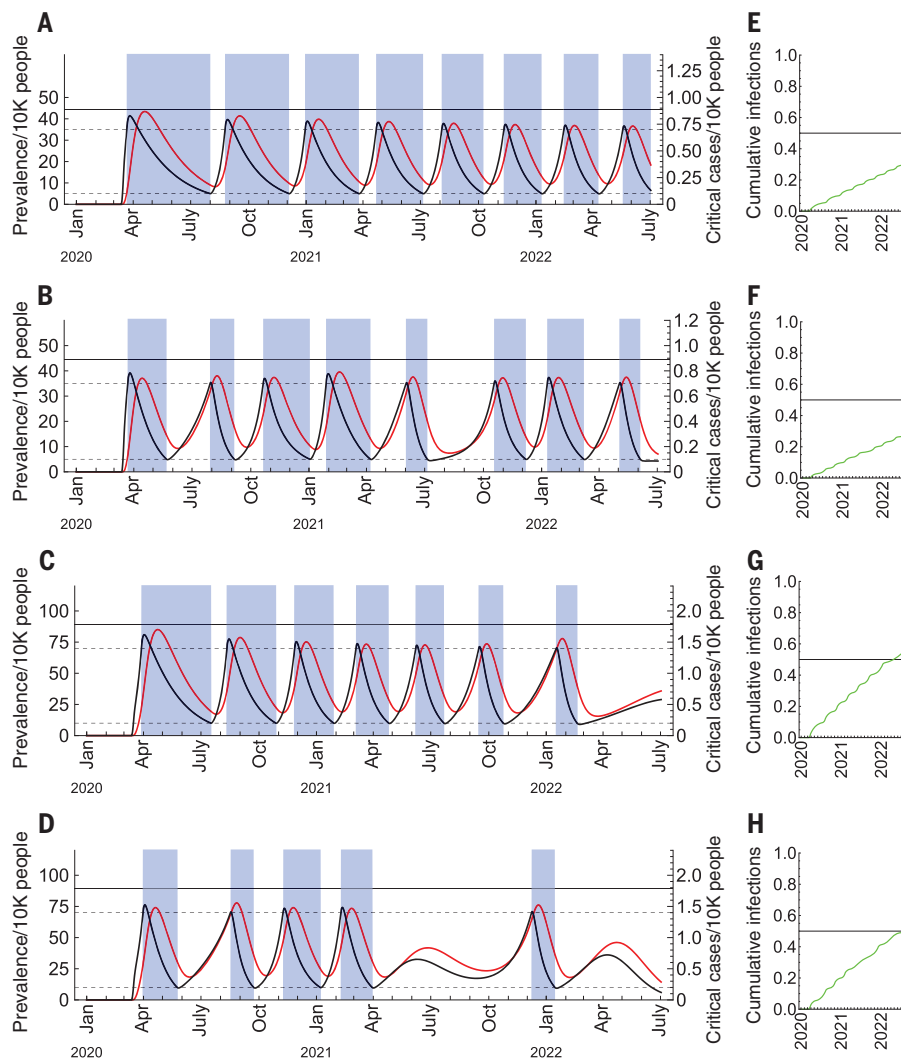


Fig. 6. Intermittent social distancing scenarios with current and expanded critical care capacity. SARS-CoV-2 prevalence (black curves) and critical cases (red curves) under intermittent social distancing (shaded blue regions) without seasonal forcing (A and C) and with seasonal forcing (B and D). Distancing yields a 60% reduction in R_0 . Critical care capacity is depicted by the solid horizontal black bars, and the on/off thresholds for social distancing are depicted by the dashed horizontal lines. (A) and (B) are the scenarios with current critical care capacity in the United States and (C) and (D) are the scenarios with double the current critical care capacity. The maximal wintertime R_0 is 2.2 and for the seasonal scenarios the summertime R_0 is 1.3 (40% decline). Prevalence is in black and critical care cases are in red. To the right of each main plot (E to H), the proportion immune over time is depicted in green with the herd immunity threshold (horizontal black bar).

may differ by age, we did not have sufficient data to parameterize an age-structured model. We also did not directly model any effect from the opening of schools, which could lead to an additional boost in transmission strength in the early autumn (45). The transmission model is deterministic, so it cannot capture the possibility of SARS-CoV-2 extinction. It also does not incorporate geographic structure, so the possibility of spatially heterogeneous transmission cannot be assessed. The construction of spatially explicit models will become more feasible as more data on SARS-CoV-2 incidence become available; these will help to determine whether there are differences in seasonal forcing between geographic locations, as is the case for influenza (12), and will also help to assess the possibility of pandemic extinction while accounting for reintroductions. The timing and strength of postpandemic outbreaks may also depend on stochastic introductions from abroad, which can be assessed using more complex, global models.

We used percent test-positive multiplied by percent ILI to approximate coronavirus incidence up to a proportional constant; results were similar when using the raw number of positive tests and the raw percent test-positive as incidence proxies (fig. S1). Although the percent test-positive multiplied by percent ILI has been shown to be one of the best available proxies for influenza incidence (32), the conversion between this measure and the true incidence of coronavirus infections is unclear, so we do not make precise estimates of the overall coronavirus incidence. This conversion will undoubtedly depend on the particular population for which these estimates are being made. In a recent study, an estimated 4% of individuals with coronavirus sought medical care, and only a fraction of these were tested (46). In addition, the method that we adopted to estimate the R_e depends on the serial interval distribution, which has not been well studied for commonly circulating human coronaviruses; we used the best-available evidence from SARS-CoV-1, the most closely related coronavirus to SARS-CoV-2.

Our findings generalize only to temperate regions, which contain 60% of the world's population (47), and the size and intensity of outbreaks could be further modulated by differences in average interpersonal contact rates by location and the timing and effectiveness of nonpharmaceutical and pharmaceutical interventions. The transmission dynamics of respiratory illnesses in tropical regions can be much more complex. However, we expect that if postpandemic transmission of SARS-CoV-2 does take hold in temperate regions, there will also be continued transmission in tropical regions seeded by the seasonal outbreaks to the north and south. With such re-

seeding, long-term disappearance of any strain becomes less likely (48), but according to our model, the R_e of SARS-CoV-2 remains <1 during most of each period when that strain disappears, meaning that reseeding would shorten these disappearances only modestly.

Our findings indicate key data required to know how the current SARS-CoV-2 outbreak will unfold. Most crucially, serological studies could indicate the extent of population immunity and whether immunity wanes and at what rate. In our model, this rate is the key modulator of the total SARS-CoV-2 incidence in the coming years. Although long-lasting immunity would lead to lower overall incidence of infection, it would also complicate vaccine efficacy trials by contributing to low case numbers when those trials are conducted, as occurred with Zika virus (49). In our assessment of control measures in the initial pandemic period, we assumed that SARS-CoV-2 infection induces immunity that lasts for at least 2 years, but social distancing measures may need to be extended if SARS-CoV-2 immunity wanes more rapidly. In addition, if serological data reveal the existence of many undocumented asymptomatic infections that lead to immunity (50), less social distancing may be required. Serology could also indicate whether cross-immunity exists among SARS-CoV-2, HCoV-OC43, and HCoV-HKU1, which could affect the postpandemic transmission of SARS-CoV-2. We anticipate that such cross-immunity would lessen the intensity of SARS-CoV-2 outbreaks, though some speculate that antibody-dependent enhancement (ADE) induced by prior coronavirus infection may increase susceptibility to SARS-CoV-2 and exacerbate the severity of infection (51, 52). At present, there are limited data describing ADE between coronaviruses, but if it does exist, it may promote the cocirculation of betacoronavirus strains.

To implement intermittent social distancing, it will be necessary to carry out widespread viral testing for surveillance to monitor when the prevalence thresholds that trigger the beginning or end of distancing have been crossed. Without such surveillance, critical care bed availability might be used as a proxy for prevalence, but this metric is far from optimal because the lag between distancing and peak critical care demand could lead to frequent overrunning of critical care resources. Critical care resources are also at greater risk of being overrun if the infectious, latent, and hospitalized periods follow peaked distributions (e.g., gamma versus exponential). Measuring the distributions of these times, and not just their means, will help to set more effective thresholds for distancing interventions. Under some circumstances, intense social distancing may be able to reduce the prevalence of COVID-19 enough to warrant a shift in

strategy to contact tracing and containment efforts, as has occurred in many parts of China (21, 23, 53). Still, countries that have achieved this level of control of the outbreak should prepare for the possibility of substantial resurgences of infection and a return to social distancing measures, especially if seasonal forcing contributes to a rise in transmissibility in the winter. Moreover, a winter peak for COVID-19 would coincide with peak influenza incidence (54), further straining health care systems.

Treatments or vaccines for SARS-CoV-2 would reduce the duration and intensity of the social distancing required to maintain control of the pandemic. Treatments could reduce the proportion of infections that require critical care and the duration of infectiousness, which would both directly and indirectly (through a reduction in R_0) reduce the demand for critical care resources. A vaccine would accelerate the accumulation of immunity in the population, reducing the overall length of the pandemic and averting infections that might have resulted in a need for critical care. Furthermore, if there have been many undocumented immunizing infections, then the herd immunity threshold may be reached sooner than our models suggest. Nevertheless, SARS-CoV-2 has demonstrated an ability to challenge robust healthcare systems, and the development and widespread adoption of pharmaceutical interventions will take months at best, so a period of sustained or intermittent social distancing will almost certainly be necessary.

In summary, the total incidence of COVID-19 illness over the next 5 years will depend critically upon whether it enters into regular circulation after the initial pandemic wave, which in turn depends primarily upon the duration of immunity that SARS-CoV-2 infection imparts. The intensity and timing of pandemic and postpandemic outbreaks will depend on the time of year when widespread SARS-CoV-2 infection becomes established and, to a lesser degree, upon the magnitude of seasonal variation in transmissibility and the level of cross-immunity that exists between the betacoronaviruses. Social distancing strategies could reduce the extent to which SARS-CoV-2 infections strain health care systems. Highly effective distancing could reduce SARS-CoV-2 incidence enough to make a strategy that is based on contact tracing and quarantine feasible, as in South Korea and Singapore. Less effective one-time distancing efforts may result in a prolonged single-peak pandemic, with the extent of strain on the healthcare system and the required duration of distancing depending on the effectiveness. Intermittent distancing may be required into 2022 unless critical care capacity is increased substantially or a treatment or vaccine becomes available.

The authors are aware that prolonged distancing, even if intermittent, is likely to have profoundly negative economic, social, and educational consequences. Our goal in modeling such policies is not to endorse them, but rather to identify likely trajectories of the pandemic under alternative approaches, to identify complementary interventions such as expanding ICU capacity and identifying treatments to reduce ICU demand, and to spur innovative ideas (55) to expand the list of options to bring the pandemic under long-term control. Our model presents a variety of scenarios intended to anticipate possible SARS-CoV-2 transmission dynamics under specific assumptions. We do not take a position on the advisability of these scenarios given the economic burden that sustained distancing may impose, but we note the potentially catastrophic burden on the healthcare system that is predicted if distancing is poorly effective and/or not sustained for long enough. The model will have to be tailored to local conditions and updated as more accurate data become available. Longitudinal serological studies are urgently required to determine the extent and duration of immunity to SARS-CoV-2, and epidemiological surveillance should be maintained in the coming years to anticipate the possibility of resurgence.

REFERENCES AND NOTES

- World Health Organization, *Coronavirus Disease 2019 (COVID-19) Situation Report – 66* (WHO, 2020); https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200326-sitrep-66-covid-19.pdf?sfvrsn=9e5b8b48_2.
- R. Li, C. Rivers, Q. Tan, M. B. Murray, E. Toner, M. Lipsitch, The demand for inpatient and ICU beds for COVID-19 in the US: lessons from Chinese cities. *medRxiv* 2020.03.09.20033241 [Preprint]. 16 March 2020; <https://doi.org/10.1101/2020.03.09.20033241>.
- "Not a wave, a tsunami: Italy hospitals at virus limit," *AP NEWS*, 13 March 2020; <https://apnews.com/a4497f31bf5dbc1ff263e4263fc9f69e>.
- "COVID-19 infections rise in New York with peak weeks away," *AP NEWS*, 25 March 2020; <https://apnews.com/7c7563cb82626a4024797c6aa6da260a>.
- D. L. Heymann, J. S. Mackenzie, M. Peiris, *Lancet* **381**, 779–781 (2013).
- Centers for Disease Control and Prevention, *Transcript for the CDC Telebriefing Update on COVID-19* (CDC, 2020); <https://www.cdc.gov/media/releases/2020/t0225-cdc-telebriefing-covid-19.html>.
- P. S. Wikramaratna, M. Sandeman, M. Recker, S. Gupta, *Philos. Trans. R. Soc. B Biol. Sci.* **368**, 20120200 (2013).
- L. Vijgen et al., *J. Virol.* **79**, 1595–1604 (2005).
- S. Su et al., *Trends Microbiol.* **24**, 490–502 (2016).
- M. E. Killerby et al., *J. Clin. Virol.* **101**, 52–56 (2018).
- R. A. Neher, R. Dyrda, V. Druelle, E. B. Hodcroft, J. Albert, *Swiss Med. Wkly.* **150**, w20224 (2020).
- J. Shaman, V. E. Pitzer, C. Viboud, B. T. Grenfell, M. Lipsitch, *PLOS Biol.* **8**, e1000316 (2010).
- J. Shaman, E. Goldstein, M. Lipsitch, *Am. J. Epidemiol.* **173**, 127–135 (2011).
- I. Chattopadhyay, E. Kiciman, J. W. Elliott, J. L. Shaman, A. Rzhetsky, *eLife* **7**, e30756 (2018).
- K. A. Callow, H. F. Parry, M. Sergeant, D. A. Tyrrell, *Epidemiol. Infect.* **105**, 435–446 (1990).
- K.-H. Chan et al., *J. Infect.* **67**, 130–140 (2013).
- D. M. Patrick et al., *Can. J. Infect. Dis. Med. Microbiol.* **17**, 330–336 (2006).
- Z. Wu, J. M. McGoogan, *JAMA* **323**, 1239 (2020).
- A. Hauser, M. J. Counotte, C. C. Margossian, G. Konstantinoudis, N. Low, C. L. Althaus, J. Riou, Estimation of SARS-CoV-2 mortality during the early stages of an epidemic: a modelling study in Hubei, China and northern Italy. *medRxiv* 2020.03.04.20031104 [Preprint]. 6 March 2020; <https://doi.org/10.1101/2020.03.04.20031104>.
- R. Verity, L. C. Okell, I. Dorigatti, P. Winskill, C. Whittaker, N. Imai, G. Cuomo-Dannenburg, H. Thompson, P. Walker, H. Fu, A. Dighe, J. Griffin, A. Cori, M. Baguelin, S. Bhatia, A. Boonyasiri, Z. M. Cucunubá, R. FitzJohn, K. A. M. Gaythorpe, W. Green, A. Hamlet, W. Hinsley, D. Laydon, G. Nedjati-Gilani, S. Riley, S. van Elsland, E. Volz, H. Wang, Y. Wang, X. Xi, C. Donnelly, A. Ghani, N. Ferguson, Estimates of the severity of COVID-19 disease. *medRxiv* 2020.03.09.20033357 [Preprint]. 13 March 2020; <https://doi.org/10.1101/2020.03.09.20033357>.
- L. Ferretti et al., *Science* eabb6936 (2020).
- R. M. Anderson, H. Heesterbeek, D. Klinkenberg, T. D. Hollingsworth, *Lancet* **395**, 931–934 (2020).
- Q. Bi, Y. Wu, S. Mei, C. Ye, X. Zou, Z. Zhang, X. Liu, L. Wei, S. A. Truelove, T. Zhang, W. Gao, C. Cheng, X. Tang, X. Wu, Y. Wu, B. Sun, S. Huang, Y. Sun, J. Zhang, T. Ma, J. Lessler, T. Feng, Epidemiology and Transmission of COVID-19 in Shenzhen China: Analysis of 391 cases and 1,286 of their close contacts. *medRxiv* 2020.03.03.20028423 [Preprint]. 27 March 2020; <https://doi.org/10.1101/2020.03.03.20028423>.
- N. Thakkar, R. Burstein, H. Hu, P. Selvaraj, D. Klein, Institute for Disease Modeling, Bill & Melinda Gates Foundation, Social distancing and mobility reductions have reduced COVID-19 transmission in King County, WA (Institute for Disease Modeling, 2020); https://covid.idmod.org/data/Social_distancing_mobility_reductions_reduced_COVID_Seattle.pdf.
- S. Lai, N. W. Ruktanonchai, L. Zhou, O. Prosper, W. Luo, J. R. Floyd, A. Wesolowski, M. Santillana, C. Zhang, X. Du, H. Yu, A. J. Tatem, Effect of non-pharmaceutical interventions for containing the COVID-19 outbreak in China. *medRxiv* 10.1101/2020.03.03.20029843 [Preprint]. 13 March 2020; <https://doi.org/10.1101/2020.03.03.20029843>.
- N. M. Ferguson, D. Laydon, G. Nedjati-Gilani, N. Imai, K. Ainslie, M. Baguelin, S. Bhatia, A. Boonyasiri, Z. Cucunubá, G. Cuomo-Dannenburg, A. Dighe, H. Fu, K. Gaythorpe, H. Thompson, R. Verity, E. Volz, H. Wang, Y. Wang, P. G. Walker, C. Walters, P. Winskill, C. Whittaker, C. A. Donnelly, S. Riley, A. C. Ghani, Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand (Imperial College COVID-19 Response Team, 2020); <https://www.imperial.ac.uk/media/imperial-college/medicine/sph/ide/gida-fellowships/Imperial-College-COVID19-NPI-modelling-16-03-2020.pdf>.
- "Coronavirus: Thousands of extra hospital beds and staff," *BBC News*, 21 March 2020; <https://www.bbc.com/news/uk-51989183>.
- C. Chen, B. Zhao, *J. Hosp. Infect.* S0195-6701(20)30107-9 (2020).
- "Pentagon says it will give 5 million respirators, 2,000 ventilators to Health and Human Services for virus response," *AP NEWS*, 17 March 2020; <https://apnews.com/79e98812b5b1592a134803b00c8d88b0>.
- "Coronavirus: How can China build a hospital so quickly?" *BBC News*, 31 January 2020; <https://www.bbc.com/news/world-asia-china-51245156>.
- Centers for Disease Control and Prevention, *The National Respiratory and Enteric Virus Surveillance System (NREVSS)* (CDC, 2020); <https://www.cdc.gov/surveillance/nrevss/index.html>.
- E. Goldstein, S. Cobey, S. Takahashi, J. C. Miller, M. Lipsitch, *PLOS Med.* **8**, e1001051 (2011).
- Centers for Disease Control and Prevention, *FluView Interactive* (CDC, 2018); <https://www.cdc.gov/flu/weekly/fluviewinteractive.htm>.
- J. Wallinga, P. Teunis, *Am. J. Epidemiol.* **160**, 509–516 (2004).
- J. Wallinga, M. Lipsitch, *Proc. Biol. Sci.* **274**, 599–604 (2007).
- D. E. te Beest, M. van Boven, M. Hooiveld, C. van den Dool, J. Wallinga, *Am. J. Epidemiol.* **178**, 1469–1477 (2013).
- J. M. Read, J. R. Bridgen, D. A. Cummings, A. Ho, C. P. Jewell, Novel coronavirus 2019-nCoV: early estimation of epidemiological parameters and epidemic predictions. *medRxiv* 10.1101/2020.01.23.20018549 [Preprint]. 28 January 2020; <https://doi.org/10.1101/2020.01.23.20018549>.
- S. A. Lauer et al., *Ann. Intern. Med.* (2020).
- N. M. Linton et al., *J. Clin. Med.* **9**, 538 (2020).
- World Health Organization, *Coronavirus Disease 2019 (COVID-19) Situation Report – 51* (WHO, 2020); https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200311-sitrep-51-covid-19.pdf?sfvrsn=1ba62e57_10.
- Q. Li et al., *N. Engl. J. Med.* **382**, 1199–1207 (2020).
- A. Handel, I. M. Longini Jr., R. Antia, *Proc. Biol. Sci.* **274**, 833–837 (2007).
- C. M. Peak, R. Kahn, Y. H. Grad, L. M. Childs, R. Li, M. Lipsitch, C. O. Buckee, Modeling the comparative impact of individual quarantine vs. active monitoring of contacts for the mitigation of COVID-19. *medRxiv* 10.1101/2020.03.05.20031088 [Preprint]. 8 March 2020; <https://doi.org/10.1101/2020.03.05.20031088>.
- R. J. Hatchett, C. E. Mecher, M. Lipsitch, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 7582–7587 (2007).
- K. E. Huang, M. Lipsitch, J. Shaman, E. Goldstein, *Epidemiology* **25**, 203–206 (2014).
- J. Shaman, M. Galanti, Direct measurement of rates of asymptomatic infection and clinical care-seeking for seasonal coronavirus. *medRxiv* 10.1101/2020.01.30.20019612 [Preprint]. 3 February 2020; <https://doi.org/10.1101/2020.01.30.20019612>.
- James Cook University, *State of the Tropics 2017 Report: Sustainable Infrastructure in the Tropics* (James Cook University, 2017); <https://www.jcu.edu.au/state-of-the-tropics/publications/2017>.
- D. He et al., *Sci. Rep.* **5**, 11013 (2015).
- K. S. Vannice et al., *Vaccine* **37**, 863–868 (2019).
- R. Li et al., *Science* eabb3221 (2020).
- J. A. Tetro, *Microbes Infect.* **22**, 72–73 (2020).
- Y. Fu, Y. Cheng, Y. Wu, *Virol. Sin.* (2020).
- M. U. G. Kraemer et al., *Science* **368**, eabb4218 (2020).
- M. Lipsitch, C. Viboud, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 3645–3646 (2009).
- H. V. Fineberg, *N. Engl. J. Med.* (2020).
- C. Tedijanto, c2-d2/CoV-seasonality: First release, Zenodo (2020); https://zenodo.org/record/3726085#XpC68_7twwk.
- S. Kissler, nCoV_introduction, Version 4.0, Zenodo (2020); https://zenodo.org/record/3745557#XpC7Y_7twwk.

ACKNOWLEDGMENTS

We thank M. Killerby and A. Haynes for their helpful comments on early versions of this manuscript, M. W. Shelley for advice on structuring the manuscript, and B. Gnanngon for helpful discussions on viral immunity dynamics. **Funding:** C.T. was supported by T32AI007535 from the National Institute of Allergy and Infectious Diseases. The work was also supported by the Morris-Singer Fund for the Center for Communicable Disease Dynamics at the Harvard T.H. Chan School of Public Health. Y.G. was funded by the NIH, the Doris Duke Charitable Foundation, the Wellcome Trust, and internal awards through Harvard University. **Author contributions:** S.M.K. conceived of the study, conducted the analysis, and wrote the manuscript. C.T. conceived of the study, conducted the analysis, and wrote the manuscript. E.G. assisted with the analysis and edited the manuscript. Y.H.G. conceived of the study, edited the manuscript, and oversaw the work. M.L. conceived of the study, edited the manuscript, and oversaw the work. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** ILINet data are publicly available through the FluView interactive website (33). Regression (56) and transmission model (57) code are available in online repositories. This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>. This license does not apply to figures/photos/artwork or other content included in the article that is credited to a third party; obtain authorization from the rights holder before using such material.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6493/860/suppl/DC1
Materials and Methods
Figs. S1 to S17
Tables S1 to S8
References (58–62)
MDAR Reproducibility Checklist

[View/request a protocol for this paper from Bio-protocol.](#)

4 March 2020; accepted 9 April 2020
Published online 14 April 2020
10.1126/science.abb5793

TROPICAL FOREST

Long-term thermal sensitivity of Earth's tropical forests

Martin J. P. Sullivan^{1,2*}, Simon L. Lewis^{1,3}, Kofi Affum-Baffoe⁴, Carolina Castilho⁵, Flávia Costa⁶, Aida Cuni Sanchez^{7,8}, Corneille E. N. Ewango^{9,10,11}, Wannes Hubau^{12,13}, Beatriz Marimon¹⁴, Abel Monteagudo-Mendoza¹⁵, Lan Qie¹⁶, Bonaventure Sonké¹⁷, Rodolfo Vasquez Martinez¹⁵, Timothy R. Baker¹, Roel J. W. Brienen¹, Ted R. Feldpausch¹⁸, David Galbraith¹, Manuel Gloor¹, Yadvinder Malhi¹⁹, Shin-ichiro Aiba²⁰, Miguel N. Alexiades²¹, Everton C. Almeida²², Edmar Almeida de Oliveira²³, Esteban Álvarez Dávila²⁴, Patricia Alvarez Loayza²⁵, Ana Andrade²⁶, Simone Aparecida Vieira²⁷, Luiz E. O. C. Aragão^{18,28}, Alejandro Araujo-Murakami²⁹, Eric J. M. M. Arets³⁰, Luzmila Arroyo³¹, Peter Ashton³², Gerardo Aymard C.³³, Fabrício B. Baccaro³⁴, Lindsay F. Banin³⁵, Christopher Baraloto³⁶, Plínio Barbosa Camargo³⁷, Jos Barlow³⁸, Jorcely Barroso³⁹, Jean-François Bastin^{40,41}, Sarah A. Batterman^{1,42,43,44}, Hans Beeckman¹², Serge K. Begne^{1,17}, Amy C. Bennett¹, Erika Berenguer^{19,38}, Nicholas Berry⁴⁵, Lilian Blanc⁴⁶, Pascal Boeckx⁴⁷, Jan Bogaert⁴⁸, Damien Bonal⁴⁹, Frans Bongers⁵⁰, Matt Bradford⁵¹, Francis Q. Brearley², Terry Brncic⁵², Foster Brown⁵³, Benoit Burban⁵⁴, José Luís Camargo²⁶, Wendeson Castro⁵⁵, Carlos Céron⁵⁶, Sabina Cerruto Ribeiro⁵⁷, Victor Chama Moscoso¹⁵, Jérôme Chave⁵⁸, Eric Chezeaux⁵⁹, Connie J. Clark²⁵, Fernanda Coelho de Souza¹, Murray Collins^{60,61}, James A. Comiskey^{62,63}, Fernando Cornejo Valverde⁶⁴, Massiel Corrales Medina⁶⁵, Lola da Costa⁶⁶, Martin Dančák⁶⁷, Greta C. Dargie¹, Stuart Davies⁶⁸, Nallaret Davila Cardozo⁶⁹, Thales de Haulleville^{12,48}, Marcelo Brilhante de Medeiros⁷⁰, Jhon del Aguila Pasquel⁷¹, Géraldine Derroire⁷², Anthony Di Fiore⁷³, Jean-Louis Doucet⁷⁴, Aurélie Dourdain⁷², Vincent Droissant⁷⁵, Luisa Fernanda Duque⁷⁶, Romeo Ekoungoulou⁷⁷, Fernando Elias⁷⁸, Terry Erwin⁷⁹, Adriane Esquivel-Muelbert⁸⁰, Sophie Fauset⁸¹, Joice Ferreira⁸², Gerardo Flores Llampazo⁸³, Ernest Foli⁸⁴, Andrew Ford⁵¹, Martin Gilpin¹, Jefferson S. Hall⁸⁵, Keith C. Hamer⁸⁶, Alan C. Hamilton⁸⁷, David J. Harris⁸⁸, Terese B. Hart^{89,90}, Radim Hédli^{91,92}, Bruno Herault^{47,93,94}, Rafael Herrera⁹⁵, Niro Higuchi⁶, Annette Hladik⁹⁶, Euridice Honorio Coronado⁷¹, Isau Huamantupa-Chuquimaco⁹⁷, Walter Huaraca Huasco⁹⁷, Kathryn J. Jeffery⁹⁸, Eliana Jimenez-Rojas⁹⁹, Michelle Kalamandeen^{1,100,101}, Marie Noël Kamdem Djuikouo^{11,13,17,102}, Elizabeth Kearsley⁴¹, Ricardo Keichi Umetsu¹⁰³, Lip Khoon Kho¹⁰⁴, Timothy Killeen¹⁰⁵, Kanehiro Kitayama¹⁰⁶, Bente Klitgaard¹⁰⁷, Alexander Koch¹⁰⁸, Nicolas Labrière⁵⁸, William Laurance¹⁰⁹, Susan Laurance¹⁰⁹, Miguel E. Leal¹¹⁰, Aurora Levesley¹, Adriano J. N. Lima⁶, Janvier Lisingo¹¹, Aline P. Lopes²⁸, Gabriela Lopez-Gonzalez¹, Tom Lovejoy¹¹¹, Jon C. Lovett^{1,107}, Richard Lowe¹¹², William E. Magnusson¹¹³, Jagoba Malumbres-Olarte^{114,115}, Ângelo Gilberto Manzatto¹¹⁶, Ben Hur Marimon Jr.¹¹⁷, Andrew R. Marshall^{8,118,119}, Toby Marthews¹²⁰, Simone Matias de Almeida Reis^{14,19}, Colin Maycock¹²¹, Karina Melgaço¹, Casimiro Mendoza¹²², Faizah Metali¹²³, Vianet Mihindou^{124,125}, William Milliken¹⁰⁷, Edward T. A. Mitchard⁶¹, Paulo S. Morandi¹⁴, Hannah L. Mossman², Laszlo Nagy¹²⁶, Henrique Nascimento⁶, David Neill¹²⁷, Reuben Nilus¹²⁸, Percy Núñez Vargas⁹⁵, Walter Palacios¹²⁹, Nadir Pallqui Camacho^{1,95}, Julie Peacock¹, Colin Pendry⁸⁸, Maria Cristina Peñuela Mora¹³⁰, Georgia C. Pickavance¹, John Pipoly¹³¹, Nigel Pitman¹³², Maureen Playfair¹³³, Lourens Poorter⁵⁰, John R. Poulsen²⁵, Axel Dalberg Poulsen⁸⁸, Richard Preziosi², Adriana Prieto¹³⁴, Richard B. Primack¹³⁵, Hirma Ramírez-Angulo¹³⁶, Jan Reitsma¹³⁷, Maxime Réjou-Méchain⁷⁵, Zorayda Restrepo Correa¹³⁸, Thaiane Rodrigues de Sousa⁶, Lily Rodriguez Bayona¹³⁹, Anand Roopsind¹⁴⁰, Agustín Rudas¹³⁴, Ervan Rutishauser^{43,141}, Kamariah Abu Salim¹²³, Rafael P. Salomão^{142,143}, Juliana Schietti⁶, Douglas Sheil¹⁴⁴, Richarlly C. Silva^{57,145}, Javier Silva Espejo¹⁴⁶, Camila Silva Valeria³⁸, Marcos Silveira⁵⁷, Murielle Simo-Droissart¹⁷, Marcelo Fragomeni Simon⁷⁰, James Singh¹⁴⁷, Yahn Carlos Soto Shareva¹⁵, Clement Stahl⁵⁴, Juliana Stropp¹⁴⁸, Rahayu Sukri¹²³, Terry Sunderland^{149,150}, Martin Svátek¹⁵¹, Michael D. Swaine¹⁵², Varun Swamy¹⁵³, Hermann Taedoum^{154,155}, Joey Talbot¹⁵⁶, James Taplin¹⁵⁶, David Taylor¹⁵⁷, Hans ter Steege^{158,159}, John Terborgh²⁵, Raquel Thomas¹⁴⁰, Sean C. Thomas¹⁶⁰, Armando Torres-Lezama¹⁶¹, Peter Umuay^{162,163}, Luis Valenzuela Gamarra¹⁵, Geertje van der Heijden¹⁶⁴, Peter van der Hout¹⁶⁵, Peter van der Meer¹⁶⁶, Mark van Nieuwstadt¹⁶⁷, Hans Verbeeck⁴¹, Ronald Vernimmen¹⁶⁸, Alberto Vicentini⁶, Ima Célia Guimarães Vieira¹⁴⁵, Emilio Vilanova Torre¹⁶⁹, Jason Vleminckx³⁶, Vincent Vos^{170,171}, Ophelia Wang¹⁷², Lee J. T. White^{98,124,173}, Simon Willcock¹⁷⁴, John T. Woods¹⁷⁵, Verginia Wortel¹⁷⁶, Kenneth Young¹⁷⁷, Roderick Zagt¹⁷⁸, Lise Zemagho¹⁷, Pieter A. Zuidema⁵⁰, Joeri A. Zwerts^{133,167}, Oliver L. Phillips¹

The sensitivity of tropical forest carbon to climate is a key uncertainty in predicting global climate change. Although short-term drying and warming are known to affect forests, it is unknown if such effects translate into long-term responses. Here, we analyze 590 permanent plots measured across the tropics to derive the equilibrium climate controls on forest carbon. Maximum temperature is the most important predictor of aboveground biomass (−9.1 megagrams of carbon per hectare per degree Celsius), primarily by reducing woody productivity, and has a greater impact per °C in the hottest forests (>32.2°C). Our results nevertheless reveal greater thermal resilience than observations of short-term variation imply. To realize the long-term climate adaptation potential of tropical forests requires both protecting them and stabilizing Earth's climate.

The response of tropical terrestrial carbon to environmental change is a critical component of global climate models (1). Land-atmosphere feedbacks depend on the balance of positive biomass growth stimulation by CO₂ fertilization (i.e., β) and negative responses to warmer temperatures and any change in precipitation (i.e., γ). Yet the climate response is so poorly constrained that it remains one of the largest uncertainties in Earth system models (2, 3), with the temperature sensitivity of tropical land carbon

stocks alone differing by >100 Pg C °C^{−1} among models (2). Such uncertainty impedes our understanding of the global carbon cycle, limiting our ability to simulate the future of the Earth system under different long-term climate mitigation strategies. A critical long-term control on tropical land-atmosphere feedbacks is the sensitivity to climate of tropical forests (a key component of γ), where about 40% of the world's vegetation carbon resides (4).

The sensitivity to environmental change of tropical biomass carbon stocks, rates of production, and the persistence of fixed carbon can all be estimated by relating their short-term and interannual responses to variation in climate (5–7). These sensitivities are then used to con-

strain longer-term projections of climate responses (2). Such approaches typically find that higher minimum temperatures are strongly associated with slower tree growth and reduced forest carbon stocks, likely owing to increased respiration at higher temperatures (7–9). Tropical forest carbon is also sensitive to precipitation (10), with, for example, increased tree mortality occurring during drought events (11).

Yet the sensitivity of ecosystems to interannual fluctuations may be an unreliable guide to their longer-term responses to climate change. Such responses will also be influenced by physiological acclimation (12), changes in demographic rates (13), and shifts in species composition (14). For example, both respiration

Author affiliations are listed at the end of this paper.

*Corresponding author. Email: m.j.sullivan@leeds.ac.uk

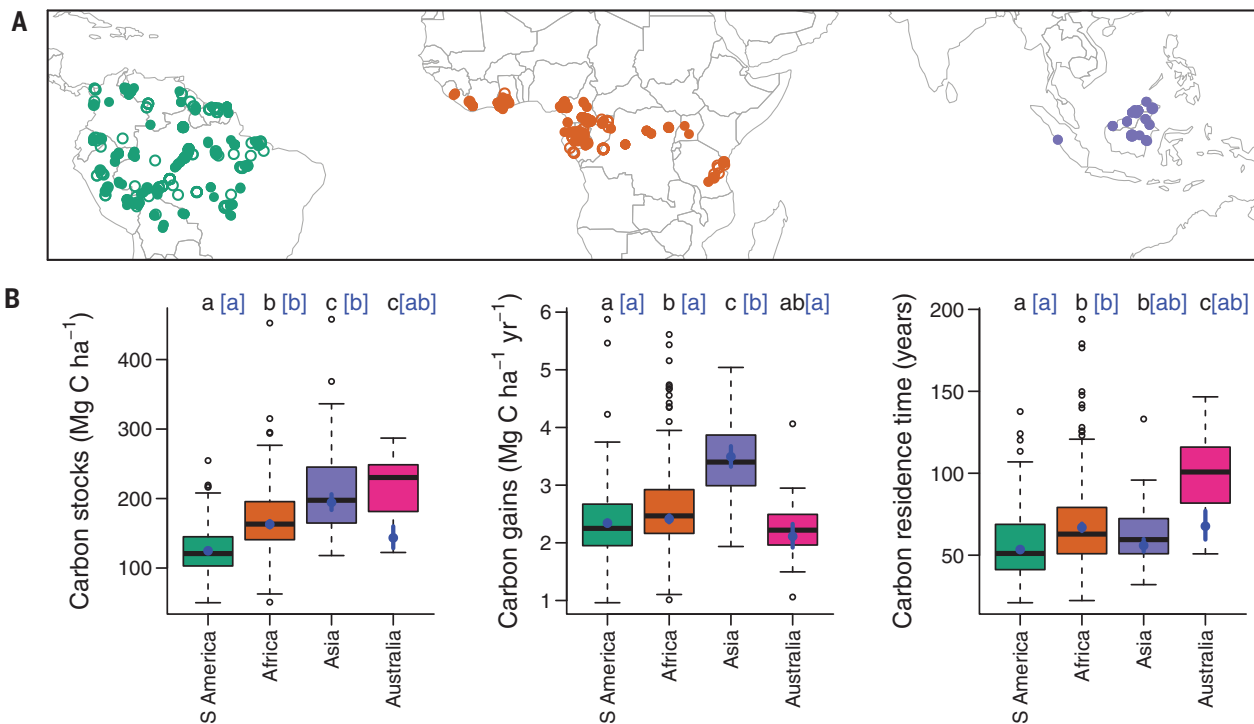


Fig. 1. Spatial variation in tropical forest carbon. (A) The RAINFOR (South America), AfriTRON (Africa), T-FORCES (Asia), and Australian plot networks. Filled symbols show 590 multicensus plots used in the main analysis; open symbols show 223 single-census plots used as an independent dataset. Symbol color indicates the region: green, South America; orange, Africa; purple, Asia; and pink, Australia. (B) Variation in carbon among continents. Boxplots show raw variation, whereas blue points show estimated mean values (\pm SE) after accounting for environmental variation. Letters denote statistically significant differences between continents ($P < 0.05$) based on raw data (black) or after accounting for environmental effects (blue in brackets).

and photosynthesis can acclimate under sustained temperature increases (15–17), tropical trees exhibit physiological plasticity (18), and shifts in species composition occur (14) under sustained drought. These processes could mean that tropical forests are less sensitive to climate than estimates derived from interannual variability imply. An alternative, complementary approach to assessing sensitivity to climate is to measure and analyze spatial variation in tropical ecosystems across climate gradients as a space-for-time substitution. Such biome-wide spatial variation in forest carbon stocks, fluxes, and persistence offers a distinctive and largely unexplored window into the potential equilibrium sensitivity of tropical forest vegetation to warming, because it captures real-world vegetation responses that allow for physiological and ecological adaptation (12).

To assess the long-term climate controls on tropical forest growth and carbon stocks, we assembled, measured, and analyzed a pantropical network of 590 permanent, long-term inventory plots (Fig. 1; see figs. S1 and S2 for ability to capture biome climate space). Our analysis combines standardized measurements from across South American, African, Asian, and Australian tropical lowland forests (273, 239, 61, and 17 plots, respectively). For every plot, we calculated aboveground carbon stocks

(19). Then, to better assess the dynamic controls on aboveground carbon stocks, we also computed the rate of carbon gained by the system (aboveground woody carbon production, calculated as tree growth plus newly recruited trees, in Mg C ha⁻¹ year⁻¹) and the carbon residence time in living biomass (calculated as the ratio of living carbon stocks to carbon gains, in years).

We found considerable variation in biomass carbon among continents, with lower stocks per unit area in South America compared with the Paleotropics, even after accounting for environmental variables (Fig. 1). Continents with high carbon stocks had either large carbon gains (Asia) or long carbon residence times (Africa) (Fig. 1). Because of these differences among continents, which are potentially due to differences in evolutionary history (20), we analyzed the environmental drivers of spatial variation in carbon stocks while accounting for biogeographical differences. We fitted linear models with explanatory variables representing hypothesized mechanistic controls of climate on tropical forest carbon (table S1). We also included soil covariates, continent intercepts, and eigenvectors describing spatial relationships among plots to account for other sources of variation (21).

Forest carbon stocks were most strongly related to maximum temperature [Fig. 2; -5.9% per 1°C increase in mean daily maximum tem-

perature in the warmest month with a 95% confidence interval (CI) = -8.6 to -3.1% , which is equivalent to -9.1 Mg C ha⁻¹ °C⁻¹ for a stand with the mean carbon stock in our dataset, 154.6 Mg C ha⁻¹] followed by rainfall (Fig. 2; $+2.4\%$ per 100-mm increase in precipitation in the driest quarter with a 95% CI = 0.6 to 4.3%, equivalent to 0.04 Mg C ha⁻¹ mm⁻¹ for a stand with the mean carbon stocks in our dataset), with no statistically significant relationship with minimum temperature, wind speed, or cloud cover (Fig. 2). The effects of maximum temperature and precipitation are also evident in an analysis considering a wider suite of climate variables than those tied to hypothesized mechanisms (fig. S3) and in an additional independent pantropical dataset of 223 single-census plots (for which carbon gains and residence time cannot be assessed, fig. S4).

The negative effect of maximum temperature on aboveground carbon stocks mainly reflects reduced carbon gains with increasing temperature (-4.0% per 1°C, 95% CI = -6.2 to -1.8% ; Fig. 2), whereas the positive effect of precipitation emerges through longer carbon residence times with increasing precipitation in the driest quarter (3.3% per 100 mm, 95% CI = 0.9 to 5.7%; Fig. 2). Carbon residence time also increased with the proportion of clay in the soil (Fig. 2). The additive effects of precipitation and temperature on carbon stocks

were modified by an interaction between them [change in Akaike information criterion (ΔAIC) = 15.4 comparing the full linear model with or without interaction], with temperature effects more negative when precipitation is low (fig. S6). The interaction was through shortening carbon residence time (ΔAIC = 11.9) rather than reducing carbon gains (model without interaction performed better; ΔAIC = 1.4).

An alternative analysis using decision-tree algorithms (22) also showed maximum temperature and precipitation to be important (fig. S7). This decision-tree approach, which can capture complex nonlinear relationships (22), indicated potential nonlinearity in the relationships between carbon stocks and both temperature and precipitation, with the positive effect of increasing dry-season precipitation on residence times strengthening when precipitation was low and the negative effect of maximum temperature intensifying at high temperatures (fig. S7).

We further investigated nonlinearity in the temperature relationship using breakpoint regression (supported over linear regression based on lower AIC, ΔAIC = 15.0), which revealed that above 32.2°C (95% CI = 31.7° to 32.6°C), the relationship between carbon stocks and maximum temperature became more negative (cooler than breakpoint, $-3.8\% \text{ } ^\circ\text{C}^{-1}$, and warmer than breakpoint, $-14.7\% \text{ } ^\circ\text{C}^{-1}$; Fig. 3). By partitioning carbon stocks into their production and persistence, we found that this nonlinearity reflects changes to carbon residence time (ΔAIC = 10.6) rather than gains (ΔAIC = 1.7). Overall, our results thus indicate two separate climate controls on carbon stocks: a negative linear effect of maximum temperature through reduced carbon gains and a

nonlinear negative effect of maximum temperature, ameliorated by high dry-season precipitation, through reduced carbon residence time.

The effect of temperature on carbon residence time only emerges when dry-season precipitation is low; this is consistent with theoretical expectations that negative effects of temperature on tree longevity are exacerbated by moisture limitation, rather than being independent of it and only due to increased respiration costs (23). This could occur through high vapor pressure deficits in hot and dry forests increasing mortality risk by causing hydraulic stress (23, 24) or carbon starvation due to limited photosynthesis as a result of stomatal closure (23). Notably, the temperature-precipitation interaction we found for aboveground stocks is in the opposite direction to temperature-precipitation interactions reported for soil carbon (25). In soils, moisture limitation suppresses the temperature response of heterotrophic respiration, whereas in trees, moisture limitation increases the mortality risks of high temperatures.

The negative effects of temperature on biomass carbon stocks and gains are primarily due to maximum rather than minimum temperature. This is consistent with high daytime temperatures reducing CO_2 assimilation rates, for example, owing to increased photorespiration or longer duration of stomatal closure (26, 27), whereas if negative temperature effects were to have increased respiration rates, there should be a stronger relationship with minimum (i.e., nighttime) temperature. Critically, minimum temperature is unrelated to aboveground carbon stocks both pantropically and in one continent, South America, where maximum and minimum temperature are largely de-

coupled [correlation coefficient (r) = 0.33; fig. S8]. Although carbon gains are negatively related to minimum temperature (fig. S9), this bivariate relationship is weaker than with maximum temperature and disappears once the effects of other variables are accounted for (Fig. 2). Finally, in Asia, the tropical region that experiences the warmest minimum temperatures of all, both carbon stocks and carbon gains are highest (Fig. 1 and fig. S11).

Overall, our results suggest that tropical forests have considerable potential to acclimate and adapt to the effects of nighttime minimum temperatures but are clearly sensitive to the effects of daytime maximum temperature. This is consistent with ecophysiological observations suggesting that the acclimation potential of respiration (15) is greater than that of photosynthesis (17). The temperature sensitivity revealed by our analysis is also considerably weaker than short-term sensitivities associated with interannual climate variation (7–9). For example, by relating short-term annual climate anomalies to responses in plots, the effect of a 1°C increase in temperature on carbon gains has been estimated as more than threefold our long-term, pantropical result (28). This stronger, long-term thermal resilience is likely due to a combination of individual acclimation and plasticity (15–17), differences in species' climate responses (29) leading to shifts in community composition due to changing demographic rates (12), and the immigration of species with higher performance at high temperatures (12).

Our pantropical analysis of the sensitivity to climate of aboveground forest carbon stocks, gains, and persistence shows that warming reduces carbon stocks and woody productivity. Using a reference carbon stock map (30) and applying our estimated temperature sensitivity (including nonlinearity) while holding other variables constant leads to an eventual biome-wide reduction of 14.1 Pg C in live biomass (including scaling to estimate carbon in roots) for a 1°C increase in mean daily maximum temperature in the warmest month (95% CI = 6.9 to 20.7 Pg). This compares with a large range of projected sensitivities in coupled climate carbon cycle models that report vegetation carbon (1 to 58 Pg $^\circ\text{C}^{-1}$), although these models have not been run to equilibrium (see supplementary methods).

Our results suggest that stabilizing global surface temperatures at 2°C above preindustrial levels will cause a potential long-term biome-wide loss of 35.3 Pg C (95% CI = 20.9 to 49.0 Pg, estimates with alternative baseline biomass maps of 24.0 to 28.4 Pg; fig. S12). The greatest long-term reductions in carbon stocks are projected in South America, where baseline temperatures and future warming are both highest (Fig. 4 and fig. S13). This warming would push 71% of the biome beyond

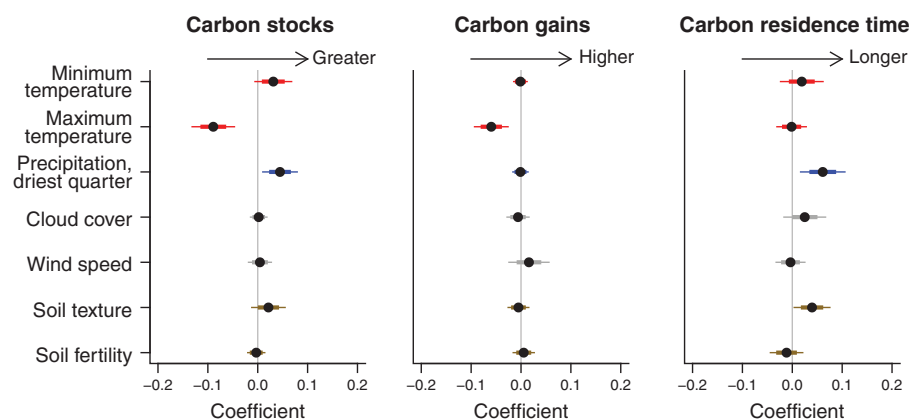


Fig. 2. Correlates of spatial variation in tropical forest carbon. Points show coefficients from model-averaged general linear models. Variables that did not occur in well-supported models are shrinkage-adjusted toward zero. Coefficients are standardized so that they represent change in the response variable for one standard deviation change in the explanatory variable. Error bars show standard errors (thick lines) and 95% confidence intervals (thin lines); error bar color is for illustrative purposes to reflect the category of variable. Soil texture is represented by the percentage clay and soil fertility by cation exchange capacity. The full models explained 44.1, 31.4, and 30.9% of spatial variation in carbon stocks, gains, and residence time, respectively. Coefficients are shown in table S2. Results are robust to using an alternative allometry to estimate tree biomass (fig. S5).

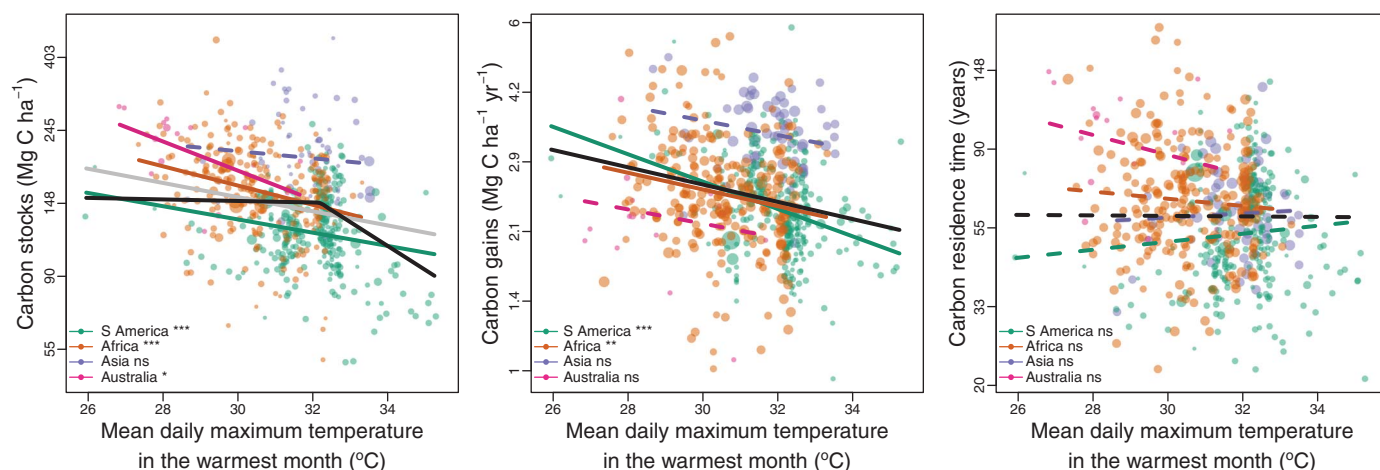


Fig. 3. Temperature effects on tropical forest carbon stocks, carbon gains, and carbon residence time. Black lines show the best pantropical relationships accounting for environmental covariates. The gray line additionally shows the linear pantropical relationship for carbon stocks. Colored lines show bivariate relationships within each continent, as identified in the legend. Statistically significant relationships are shown with solid lines; nonsignificant relationships are shown with dashed lines. The y axis is on a log scale. Symbol point size is proportional to weights used in model

fitting based on plot size and monitoring length; see supplementary materials and methods. For stocks and gains, linear and breakpoint pantropical relationships are all statistically significant ($P < 0.001$) as are better-sampled continents. For carbon residence time, relationships with temperature are nonsignificant (ns), but there is a statistically significant interaction between maximum temperature and precipitation in the driest quarter (fig. S6). Relationships with other variables are shown in figs. S8 to S10. *** $P < 0.001$; ** $P < 0.01$; * $P < 0.05$; ns, $P \geq 0.05$.

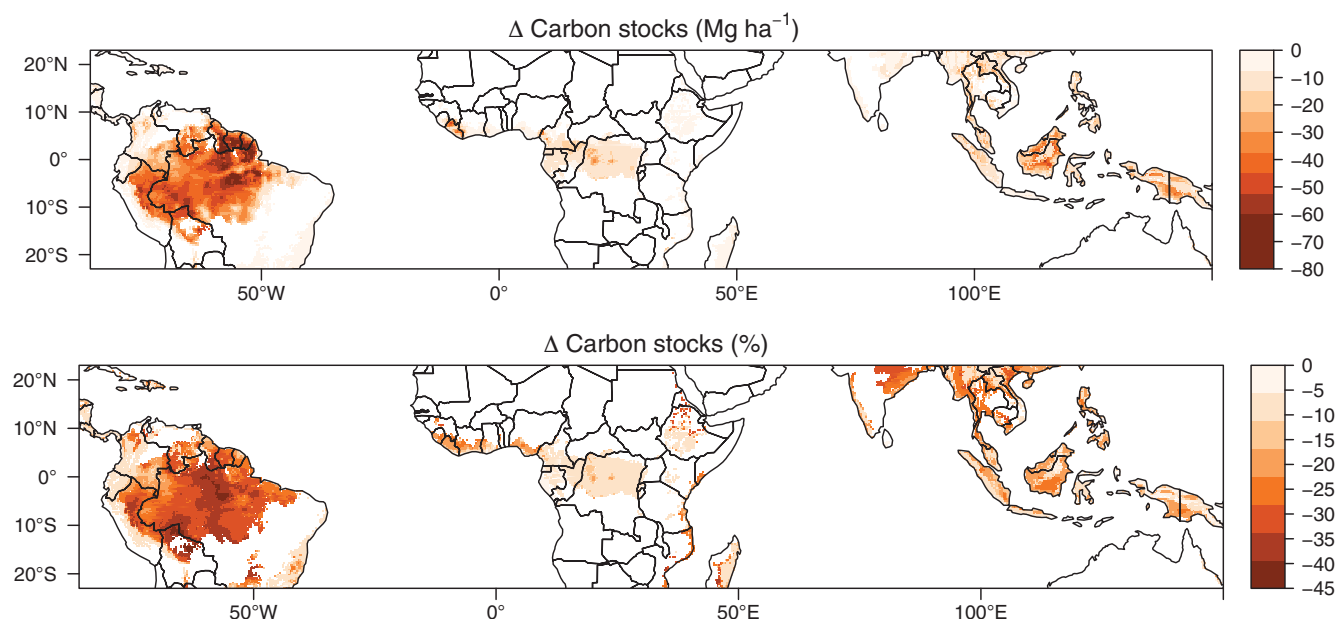


Fig. 4. Long-term change in carbon stocks due to temperature effects alone for global surface air temperature warming of 2°C. Maps show the predicted absolute and relative change in tropical forest carbon stocks. Parts of the biome become warmer than observed now in our dataset (fig. S14). See fig. S12 for predictions using alternative carbon reference maps. Predictions are based on temperature alone and do not include precipitation changes (for which future patterns of change are uncertain) or moderation by increased CO₂. (See fig. S15 for analysis incorporating this.)

the thermal threshold—a maximum temperature of 32.2°C—where larger long-term reductions in biomass are expected (fig. S14). Of course, growth stimulation by carbon dioxide (31) will partially or wholly offset the effect of this temperature increase, depending on both the level of atmospheric carbon dioxide that limits warming to 2°C above pre-industrial levels and the fertilization effect of this carbon dioxide on tropical trees. Although

CO₂ fertilization will reduce temperature-induced carbon losses from biomass across the tropics (table S3), our analysis indicates that CO₂ fertilization will not completely offset long-term temperature-induced carbon losses within Amazonia (fig. S15), consistent with a recent decadal-scale analysis of inventory data (32).

The long-term climate sensitivities derived from our pantropical field measurements

incorporate ecophysiological and ecological adaptation and so provide an estimate of the long-term, quasi-equilibrium response of tropical vegetation to climate. This thermal adaptation potential may not be fully realized in future responses because (i) the speed of temperature rises may exceed species' adaptive capabilities, (ii) habitat fragmentation may limit species' ability to track changes in the environment, and (iii) other human impacts

such as logging and fire can increase the vulnerability of forest carbon stocks to high temperatures. Although many tropical forests are under severe threat of conversion, our results show that, in the long run, tropical forests that remain intact can continue to store high levels of carbon under high temperatures. Achieving the biome-wide climate resilience potential that we document depends on limiting heating and on large-scale conservation and restoration to protect biodiversity and allow species to move.

REFERENCES AND NOTES

- P. M. Cox, R. A. Betts, C. D. Jones, S. A. Spall, I. J. Totterdell, *Nature* **408**, 184–187 (2000).
- P. M. Cox *et al.*, *Nature* **494**, 341–344 (2013).
- B. B. Booth *et al.*, *Environ. Res. Lett.* **7**, 024002 (2012).
- K.-H. Erb *et al.*, *Nature* **553**, 73–76 (2018).
- W. Wang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 13061–13066 (2013).
- J. Liu *et al.*, *Science* **358**, eaam5690 (2017).
- D. A. Clark, S. C. Piper, C. D. Keeling, D. B. Clark, *Proc. Natl. Acad. Sci. U.S.A.* **100**, 5852–5857 (2003).
- W. R. L. Anderegg *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 15591–15596 (2015).
- A. Ballantyne *et al.*, *Nat. Clim. Chang.* **7**, 148–152 (2017).
- J. K. Green *et al.*, *Nature* **565**, 476–479 (2019).
- O. L. Phillips *et al.*, *Science* **323**, 1344–1347 (2009).
- M. D. Smith, A. K. Knapp, S. L. Collins, *Ecology* **90**, 3279–3289 (2009).
- J. H. Brown, T. J. Valone, C. G. Curtin, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 9729–9733 (1997).
- S. Fauset *et al.*, *Ecol. Lett.* **15**, 1120–1129 (2012).
- C. A. Gunderson, K. H. O'Hara, C. M. Campion, A. V. Walker, N. T. Edwards, *Glob. Chang. Biol.* **16**, 2272–2286 (2010).
- M. Slot *et al.*, *Glob. Chang. Biol.* **20**, 2915–2926 (2014).
- L. F. Ow, K. L. Griffin, D. Whitehead, A. S. Walcroft, M. H. Turnbull, *New Phytol.* **178**, 123–134 (2008).
- T. F. Domingues *et al.*, *Oecologia* **187**, 933–940 (2018).
- See supplementary materials.
- J. W. F. Slik *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 1837–1842 (2018).
- S. Dray, P. Legendre, P. R. Peres-Neto, *Ecol. Modell.* **196**, 483–493 (2006).
- L. Breiman, *Mach. Learn.* **45**, 5–32 (2001).
- N. McDowell *et al.*, *New Phytol.* **219**, 851–869 (2018).
- C. G. Fontes *et al.*, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **373**, 20180209 (2018).
- P. Ciais *et al.*, *Nature* **437**, 529–533 (2005).
- M. E. Dusenage, A. G. Duarte, D. A. Way, *New Phytol.* **221**, 32–49 (2019).
- S. Pau, M. Detto, Y. Kim, C. J. Still, *Ecosphere* **9**, e02311 (2018).
- D. A. Clark, D. B. Clark, S. F. Oberbauer, *J. Geophys. Res. Biogeosci.* **118**, 783–794 (2013).
- W. R. L. Anderegg *et al.*, *Nature* **561**, 538–541 (2018).
- V. Avitabile *et al.*, *Glob. Chang. Biol.* **22**, 1406–1420 (2016).
- S. Piao *et al.*, *Glob. Chang. Biol.* **19**, 2117–2132 (2013).
- W. Hubau *et al.*, *Nature* **579**, 80–87 (2020).
- M. J. P. Sullivan, Data for “Long-term thermal sensitivity of Earth's tropical forests,” ForestPlots.NET (2020); doi: 10.5521/forestplots.net/2020_2.
- Science and Technology Development of Brazil (CNPq); Agence Nationale de la Recherche; Conservation International; Missouri Botanical Garden; Smithsonian Institution; Wildlife Conservation Society; National Geographic Society; Centre for International Forestry; and Gabon's National Park Agency. Data from RAINFOR, AfriTRON, and T-FORCES are stored and curated by ForestPlots.net, a cyber-infrastructure initiative developed at the University of Leeds that unites permanent plot records and their contributing scientists from the world's tropical forests. Many of the bodies listed above funded the development of ForestPlots.net and curation of most data analyzed here. For details of additional support and grant references, see the extended acknowledgments in the supplementary materials. **Author contributions:** M.J.P.S., O.L.P., and S.L.L. conceived, designed, and wrote the study. O.L.P., S.L.L., and Y.M. conceived the RAINFOR, AfriTRON, and T-FORCES forest census network programs. L.A., A.A.-M., T.R.B., R.J.W.B., S.K.B., K.A.-B., F.C., C.C., E.A.D., A.C.S., C.E.N.E., T.R.F., W.H., S.L.L., A.M.M., B.S.M., O.L.P., L.Q., B.S., T.S., R.V., and L.J.T.W. coordinated data collection with all authors and ForestPlots.net. M.J.P.S. analyzed the data. Full details of author contributions for data collection, management, and analysis tools are listed in the supplementary materials. All authors commented on or approved the manuscript. **Competing interests:** The authors declare no competing financial interests. **Data and materials availability:** Plot-level input data and R scripts are available as a data package on ForestPlots.net (33).
- ¹School of Geography, University of Leeds, Leeds, UK.
- ²Department of Natural Sciences, Manchester Metropolitan University, Manchester, UK. ³Department of Geography, University College London, London, UK. ⁴Mensuration Unit, Forestry Commission of Ghana, Kumasi, Ghana. ⁵Embrapa Roraima, Brazilian Agricultural Research Corporation (EMBRAPA), Brasília, Brazil. ⁶Instituto Nacional de Pesquisas da Amazônia (INPA), Manaus, Brazil. ⁷Department of Ecosystem Science and Sustainability, Colorado State University, Fort Collins, CO, USA. ⁸Department of Environment and Geography, University of York, York, UK. ⁹DR Congo Programme, Wildlife Conservation Society, Kisangani, Democratic Republic of Congo. ¹⁰Centre de Formation et de Recherche en Conservation Forestière (CEFRECOC), Epulu, Democratic Republic of Congo. ¹¹Faculté de Gestion de Ressources Naturelles Renouvelables, Université de Kisangani, Kisangani, Democratic Republic of Congo. ¹²Service of Wood Biology, Royal Museum for Central Africa, Tervuren, Belgium. ¹³Department of Environment, Laboratory of Wood Technology (Woodlab), Ghent University, Ghent, Belgium. ¹⁴UNEMAT - Universidade do Estado de Mato Grosso, Nova Xavantina-MT, Brazil. ¹⁵Jardín Botánico de Missouri, Oxapampa, Peru. ¹⁶School of Life Sciences, University of Lincoln, Lincoln, UK. ¹⁷Plant Systematics and Ecology Laboratory, Higher Teachers' Training College, University of Yaoundé I, Yaoundé, Cameroon. ¹⁸Geography, College of Life and Environmental Sciences, University of Exeter, Exeter, UK. ¹⁹Environmental Change Institute, School of Geography and the Environment, University of Oxford, Oxford, UK. ²⁰Faculty of Environmental Earth Science, Hokkaido University, Sapporo, Japan. ²¹School of Anthropology and Conservation, University of Kent, Canterbury, UK. ²²Instituto de Biodiversidade e Florestas, Universidade Federal do Oeste do Pará, Santarém - PA, Brazil. ²³Universidade do Estado de Mato Grosso, Cáceres - MT, Brazil. ²⁴Escuela de Ciencias Agrícolas, Pecuarias y del Medio Ambiente, National Open University and Distance, Bogotá, Colombia. ²⁵Nicholas School of the Environment, Duke University, Durham, NC, USA. ²⁶Projeto Dinâmica Biológica de Fragmentos Florestais, Instituto Nacional de Pesquisas da Amazônia, Manaus, Brazil. ²⁷Universidade Estadual de Campinas, Campinas, SP, Brazil. ²⁸National Institute for Space Research (INPE), São José dos Campos, SP, Brazil. ²⁹Museo de Historia Natural Noel Kempff Mercado, Universidad Autónoma Gabriel René Moreno, Santa Cruz, Bolivia. ³⁰Wageningen Environmental Research, Wageningen, Netherlands. ³¹Dirección de la Carrera de Biología, Universidad Autónoma Gabriel René Moreno, Santa Cruz, Bolivia. ³²Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA. ³³Programa de Ciencias del Agro y el Mar, Herbario Universitario, Guanare, Venezuela. ³⁴Departamento de Biología, Universidade Federal do Amazonas, Manaus, Brazil. ³⁵UK Centre of Ecology and Hydrology, Penicuik, UK. ³⁶International Center for Tropical Botany, Department of Biological Sciences, Florida International University, Miami, FL, USA. ³⁷Centro de Energia Nuclear na Agricultura, Universidade de São Paulo, São Paulo, SP, Brazil. ³⁸Lancaster Environment Centre, Lancaster University, Lancaster, UK. ³⁹Centro Multidisciplinar, Universidade Federal do Acre, Cruzeiro do Sul, AC, Brazil. ⁴⁰Institute of Integrative Biology, ETH Zurich, Zurich, Switzerland. ⁴¹Department of Environment, Computational and Applied Vegetation Ecology (CAVELab), Ghent University, Ghent, Belgium. ⁴²Priestley International Centre for Climate, University of Leeds, Leeds, UK. ⁴³Smithsonian Tropical Research Institute, Panama, Panama. ⁴⁴Cary Institute of Ecosystem Studies, Millbrook, NY, USA. ⁴⁵The Landscapes and Livelihoods Group, Edinburgh, UK. ⁴⁶UR Forest and Societies, CIRAD, Montpellier, France. ⁴⁷Isotope Bioscience Laboratory (ISOFS), Ghent University, Ghent, Belgium. ⁴⁸Gembloux Agro-Bio Tech, University of Liège, Liège, Belgium. ⁴⁹UMR Silva, INRAE, Nancy, France. ⁵⁰Forest Ecology and Forest Management Group, Wageningen University, Wageningen, Netherlands. ⁵¹CSIRO Land and Water, Australia. ⁵²Congo Programme, Wildlife Conservation Society, Brazzaville, Republic of Congo. ⁵³Woods Hole Research Center, Falmouth, MA, USA. ⁵⁴INRAE, UMR EcoFoG, CNRS, CIRAD, AgroParisTech, Université des Antilles, Université de Guyane, 97310 Kourou, French Guiana. ⁵⁵Programa de Pós-Graduação Ecologia e Manejo de Recursos Naturais, Universidade Federal do Acre, Rio Branco, AC, Brazil. ⁵⁶Herbario Alfredo Paredes, Universidad Central del Ecuador, Quito, Ecuador. ⁵⁷Centro de Ciências Biológicas e da Natureza, Universidade Federal do Acre, Rio Branco, AC, Brazil. ⁵⁸Laboratoire Évolution et Diversité Biologique, UMR 5174 (CNRS/IRD/UPS), CNRS, Toulouse, France. ⁵⁹Rougier-Gabon, Libreville, Gabon. ⁶⁰Graham Research Institute on Climate Change and the Environment, London, UK. ⁶¹School of Geosciences, University of Edinburgh, Edinburgh, UK. ⁶²Inventory and Monitoring Program, National Park Service, Fredericksburg, VA, USA. ⁶³Smithsonian Institution, Washington, DC, USA. ⁶⁴Proyecto Castaña, Made de Dios, Peru. ⁶⁵Universidad Nacional de San Agustín de Arequipa, Arequipa, Peru. ⁶⁶Instituto de Geociências, Faculdade de Meteorologia, Universidade Federal do Pará, Belém, PA, Brazil. ⁶⁷Faculty of Science, Department of Ecology and Environmental Sciences, Palacký University Olomouc, Olomouc, Czech Republic. ⁶⁸Center for Tropical Forest Science, Smithsonian Tropical Research Institute, Panama, Panama. ⁶⁹Facultad de Ciencias Biológicas, Universidad Nacional de la Amazonía Peruana, Iquitos, Peru. ⁷⁰Embrapa Genetic Resources and Biotechnology, Brazilian Agricultural Research Corporation (EMBRAPA), Brasília, Brazil. ⁷¹Instituto de Investigaciones de la Amazonia Peruana, Iquitos, Peru. ⁷²Cirad, UMR EcoFoG (AgroParisTech, CNRS, INRAE, Université des Antilles, Université de Guyane), Kourou, French Guiana. ⁷³Department of Anthropology, The University of Texas at Austin, Austin, TX, USA. ⁷⁴Forest Resources Management, Gembloux Agro-Bio Tech, University of Liège, Liège, Belgium. ⁷⁵AMAP, Université de Montpellier, IRD, CNRS, CIRAD, INRAE, Montpellier, France. ⁷⁶Socioecosistemas y Cambio Climático, Fundación con Vida, Medellín, Colombia. ⁷⁷School of Forestry, Beijing Forestry University, Beijing, China. ⁷⁸Institute of Biological Sciences, Universidade Federal do Pará, Belém, PA, Brazil. ⁷⁹National Museum of Natural History, Smithsonian Institution, Washington, DC, USA. ⁸⁰School of Geography, Earth and Environmental Sciences, University of Birmingham, Birmingham, UK. ⁸¹School of Geography, Earth and Environmental Sciences, University of Plymouth, Plymouth, UK. ⁸²Embrapa Amazônia Oriental, Brazilian Agricultural Research Corporation (EMBRAPA), Brasília, Brazil. ⁸³Universidad Nacional Jorge Basadre de Grohmann (UNJBG), Tacna, Peru. ⁸⁴Forestry Research Institute of Ghana (FORIG), Kumasi, Ghana. ⁸⁵Smithsonian Institution Forest Global Earth Observatory (ForestGEO), Smithsonian Tropical Research Institute, Washington, DC, USA. ⁸⁶School of Biology, University of Leeds, Leeds, UK. ⁸⁷128 Busbridge Lane, Godalming, Surrey, UK. ⁸⁸Royal Botanic Garden Edinburgh, Edinburgh, UK. ⁸⁹Lukuru Wildlife Research Foundation, Kinshasa, Democratic Republic of Congo. ⁹⁰Division of Vertebrate Zoology, Yale Peabody Museum of Natural History, New Haven, CT, USA. ⁹¹Institute of Botany, Czech Academy of Sciences, Brno, Czech Republic. ⁹²Department of Botany, Palacký University in Olomouc, Olomouc, Czech Republic. ⁹³CIRAD, UPR Forêts et Sociétés, Yamoussoukro, Côte d'Ivoire. ⁹⁴Institut National Polytechnique Félix Houphouët-Boigny, INP-HB,

- Yamoussoukro, Côte d'Ivoire. ⁹⁵Instituto Venezolano de Investigaciones Científicas (IVIC), Caracas, Venezuela. ⁹⁶Département Hommes, Natures, Sociétés, Muséum National d'Histoire Naturel, Paris, France. ⁹⁷Universidad Nacional de San Antonio Abad del Cusco, Cusco, Peru. ⁹⁸Biological and Environmental Sciences, University of Stirling, Stirling, UK. ⁹⁹Instituto IMANI, Universidad Nacional de Colombia, Leticia, Colombia. ¹⁰⁰Living with Lakes Centre, Laurentian University, Sudbury, Canada. ¹⁰¹Department of Plant Sciences, University of Cambridge, Cambridge, UK. ¹⁰²Faculty of Science, Department of Botany and Plant Physiology, University of Buea, Buea, Cameroon. ¹⁰³PELD, Universidade do Estado de Mato Grosso, Nova Xavantina-MT, Brazil. ¹⁰⁴Tropical Peat Research Institute, Malaysian Palm Oil Board, Selangor, Malaysia. ¹⁰⁵Agteca, Santa Cruz, Bolivia. ¹⁰⁶Graduate School of Agriculture, Kyoto University, Kyoto, Japan. ¹⁰⁷Royal Botanic Gardens Kew, Richmond, London, UK. ¹⁰⁸Department of Earth Sciences, University of Hong Kong, Pok Ful Lam, Hong Kong Special Administrative Region, China. ¹⁰⁹Centre for Tropical Environmental and Sustainability Science (TESS) and College of Marine and Environmental Sciences, James Cook University, Douglas, QLD, Australia. ¹¹⁰Uganda Programme, Wildlife Conservation Society, Kampala, Uganda. ¹¹¹Environmental Science and Policy, George Mason University, Fairfax, VA, USA. ¹¹²Botany Department, University of Ibadan, Ibadan, Nigeria. ¹¹³Coordenação da Biodiversidade, Instituto Nacional de Pesquisas da Amazônia (INPA), Manaus, Brazil. ¹¹⁴cE3c – Centre for Ecology, Evolution and Environmental Changes / Azorean Biodiversity Group, Universidade dos Açores, Angra do Heroísmo, Azores, Portugal. ¹¹⁵LIBRe – Laboratory for Integrative Biodiversity Research, Finnish Museum of Natural History, University of Helsinki, Helsinki, Finland. ¹¹⁶Laboratório de Biogeoquímica Ambiental Wolfgang C. Pfeiffer, Universidade Federal de Rondônia, Porto Velho - RO, Brazil. ¹¹⁷Faculdade de Ciências Agrárias, Biológicas e Sociais Aplicadas, Universidade do Estado de Mato Grosso, Nova Xavantina-MT, Brazil. ¹¹⁸Tropical Forests and People Research Centre, University of the Sunshine Coast, Sippy Downs, QLD, Australia. ¹¹⁹Flamingo Land Ltd., North Yorkshire, UK. ¹²⁰UK Centre for Ecology and Hydrology, Wallingford, UK. ¹²¹School of International Tropical Forestry, Universiti Malaysia Sabah, Kota Kinabalu, Malaysia. ¹²²Escuela de Ciencias Forestales, Unidad Académica del Trópico, Universidad Mayor de San Simón, Sacta, Bolivia. ¹²³Faculty of Science, Universiti Brunei Darussalam, Brunei. ¹²⁴Agence Nationale des Parcs Nationaux, Libreville, Gabon. ¹²⁵Ministère de la Forêt, de la Mer, de l'Environnement, Chargé du Plan Climat, Libreville, Gabon. ¹²⁶Institute of Biology, University of Campinas, Campinas, SP, Brazil. ¹²⁷Facultad de Ingeniería Ambiental, Universidad Estatal Amazónica, Puyo, Pastaza, Ecuador. ¹²⁸Forest Research Centre, Sabah Forestry Department, Sepilok, Malaysia. ¹²⁹Carrera de Ingeniería Forestal, Universidad Técnica del Norte, Ibarra, Ecuador. ¹³⁰Grupo de Ecosistemas Tropicales y Cambio Global, Universidad Regional Amazónica Ikiam, Tena, Ecuador. ¹³¹Public Communications and Outreach Group, Parks and Recreation Division, Oakland Park, FL, USA. ¹³²Keller Science Action Center, Field Museum, Chicago, IL, USA. ¹³³Centre for Agricultural Research in Suriname (CELOS), Paramaribo, Suriname. ¹³⁴Instituto de Ciencias Naturales, Universidad Nacional de Colombia, Leticia, Colombia. ¹³⁵Department of Biology, Boston University, Boston, MA, USA. ¹³⁶Institute of Research for Forestry Development (INDEFOR), Universidad de los Andes, Mérida, Venezuela. ¹³⁷Bureau Waardenburg, Culemborg, Netherlands. ¹³⁸Socioecosistemas y Cambio Climático, Fundación Con Vida, Medellín, Colombia. ¹³⁹Centro de Conservación, Investigación y Manejo de Áreas Naturales, CIMA Cordillera Azul, Lima, Peru. ¹⁴⁰Iwokrama International Centre for Rainforest Conservation and Development, Georgetown, Guyana. ¹⁴¹Carboforexpert, Geneva, Switzerland. ¹⁴²Universidade Federal Rural da Amazônia/CAPEs, Belém, PA, Brazil. ¹⁴³Museu Paraense Emílio Goeldi, Belém, PA, Brazil. ¹⁴⁴Faculty of Environmental Sciences and Natural Resource Management, Norwegian University of Life Sciences, Ås, Norway. ¹⁴⁵Instituto Federal do Acre, Rio Branco, AC, Brazil. ¹⁴⁶Universidad de San Antonio Abad del Cusco, Cusco, Peru. ¹⁴⁷Guyana Forestry Commission, Georgetown, Guyana. ¹⁴⁸Departamento de Biogeografía y Cambio Global, Museo Nacional de Ciencias Naturales, Consejo Superior de Investigaciones Científicas (MNCN-CSIC), Madrid, Spain. ¹⁴⁹Sustainable Landscapes and Food Systems, Center for International Forestry Research, Bogor, Indonesia. ¹⁵⁰Faculty of Forestry, University of British Columbia, Vancouver, Canada. ¹⁵¹Department of Forest Botany, Dendrology and Geobiocoenology, Mendel University in Brno, Brno, Czech Republic. ¹⁵²Department of Plant and Soil Science, School of Biological Sciences, University of Aberdeen, Aberdeen, UK. ¹⁵³Institute for Conservation Research, San Diego Zoo, San Diego, CA, USA. ¹⁵⁴Department of Plant Biology, Faculty of Sciences, University of Yaounde 1, Yaoundé, Cameroon. ¹⁵⁵Bioversity International, Yaoundé, Cameroon. ¹⁵⁶UK Research and Innovation, Innovate UK, London, UK. ¹⁵⁷Department of Geography, National University of Singapore, Singapore. ¹⁵⁸Naturalis Biodiversity Center, Leiden, Netherlands. ¹⁵⁹Systems Ecology, Vrije Universiteit Amsterdam, Amsterdam, Netherlands. ¹⁶⁰Faculty of Forestry, University of Toronto, Toronto, Canada. ¹⁶¹Universidad de los Andes, Mérida, Colombia. ¹⁶²Wildlife Conservation Society, New York, NY, USA. ¹⁶³Yale School of Forestry and Environmental Studies, Yale University, New Haven, CT, USA. ¹⁶⁴School of Geography, University of Nottingham, Nottingham, UK. ¹⁶⁵Van der Hout Forestry Consulting, Rotterdam, Netherlands. ¹⁶⁶Van Hall Larenstein University of Applied Sciences, Velp, Netherlands. ¹⁶⁷Utrecht University, Utrecht, Netherlands. ¹⁶⁸Data for Sustainability, Axel, Netherlands. ¹⁶⁹School of Environmental and Forest Sciences, University of Washington, Seattle, OR, USA. ¹⁷⁰Centro de Investigación y Promoción del Campesinado, La Paz, Bolivia. ¹⁷¹Universidad Autónoma del Beni José Ballivián, Riberalta, Bolivia. ¹⁷²School of Earth Sciences and Environmental Sustainability, Northern Arizona University, Flagstaff, AZ, USA. ¹⁷³Institut de Recherche en Ecologie Tropicale, Libreville, Gabon. ¹⁷⁴School of Natural Sciences, University of Bangor, Bangor, UK. ¹⁷⁵University of Liberia, Monrovia, Liberia. ¹⁷⁶Forest Management, Centre for Agricultural Research in Suriname (CELOS), Paramaribo, Suriname. ¹⁷⁷Department of Geography and The Environment, University of Texas at Austin, Austin, TX, USA. ¹⁷⁸Tropenbos International, Wageningen, Netherlands.

†Present address: Amazonia Green Landscape Protection and Governance Programme, SOS Amazônia, 61 Pará St., Rio Branco, AC 69905-082, Brazil. ‡Present address: College of Biological and Environmental Sciences, Universidad San Francisco de Quito, Cumbayá, Ecuador. §Present address: Institute for Transport Studies, University of Leeds, Leeds, UK.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6493/869/suppl/DC1
Extended Acknowledgments
Materials and Methods
Figs. S1 to S15
Tables S1 to S3
References (34–87)

10 May 2019; accepted 5 March 2020
10.1126/science.aaw7578

DEVICE TECHNOLOGY

Precise pitch-scaling of carbon nanotube arrays within three-dimensional DNA nanotrenches

Wei Sun^{1,2,*}, Jie Shen^{1,2,†}, Zhao Zhao^{1,3,4,†}, Noel Arellano⁵, Charles Rettner⁵, Jianshi Tang⁶, Tianyang Cao¹, Zhiyu Zhou¹, Toan Ta⁵, Jason K. Streit⁷, Jeffrey A. Fagan⁷, Thomas Schaus^{1,2}, Ming Zheng⁷, Shu-Jen Han⁶, William M. Shih^{1,3,4}, Hareem T. Maune⁵, Peng Yin^{1,2,*}

Precise fabrication of semiconducting carbon nanotubes (CNTs) into densely aligned evenly spaced arrays is required for ultrascaled technology nodes. We report the precise scaling of inter-CNT pitch using a supramolecular assembly method called spatially hindered integration of nanotube electronics. Specifically, by using DNA brick crystal-based nanotrenches to align DNA-wrapped CNTs through DNA hybridization, we constructed parallel CNT arrays with a uniform pitch as small as 10.4 nanometers, at an angular deviation $<2^\circ$ and an assembly yield $>95\%$.

Although conventional transistor lithography successfully scales the channel pitch (spacing between two adjacent channels within individual transistor) of bulk materials (that is, Si), the performance drops for patterning one-dimensional (1D) semiconductors, such as carbon nanotubes (CNTs), at ultrascaled technology nodes (1, 2). The projected channel pitches [~ 10 nm or less (1)] for multichannel CNTs are smaller than the fabrication feasibility of current lithography. Alternatively, thin-film approaches (1), which use physical forces (3–6), or chemical recognition (7–9) to assemble CNTs, provide a density exceeding 500 CNTs/ μm (3). However, assembly defects, including crossing (4, 10), bundling (i.e., multiple CNTs aggregated side by side) (3), and irregular pitches (11), are widely observed in such CNT thin films.

Structural DNA nanotechnology (12, 13), in particular DNA origami (14, 15) and DNA

bricks (16, 17), can produce user-prescribed 2D or 3D objects at 2-nm feature resolution. Self-assembled DNA structures have been used to pattern diverse materials, including oxides (18, 19), graphene (20), plasmonic materials (21, 22), polymers (23), and CNTs (8, 9, 24, 25). Despite these demonstrations, unconfined surface rotation (8, 24) still limits the precise pitch scaling achieved within a DNA template. Additionally, CNT arrays assembled by using double-stranded DNAs (dsDNAs) (8) contain only a small number of CNTs per single-orientation domain (2.4 on average), less than the desired value of six CNTs (1).

By using nanotrenches based on DNA brick crystals to spatially confine the DNA hybridization-mediated CNT alignment, we developed a spatially hindered integration of nanotube electronics (SHINE) method for building evenly spaced CNT arrays (Fig. 1). DNA hybridizations between single-stranded handles within the nanotrenches and the anti-

handles (sequences complementary to the DNA handles) on CNTs compensated for the electrostatic repulsions during assembly. DNA nanotrenches also confined the orientation of individual CNTs precisely along their longitudinal axis.

Programming the DNA trench periodicity thus rationally scaled the inter-CNT pitch from 24.1 to 10.4 nm. Misaligned CNTs could not access the DNA handles and were repelled from the DNA templates by electrostatic repulsion. The pitch precision, indicative of array uniformity, improved when compared to the values for CNT thin films (11). The design for SHINE began by constructing parallel nanotrenches along the x direction (Fig. 1). The feature-repeating unit of DNA brick crystal template (17) contained 6768 base pairs. The sidewall and the bottom layer within the unit consisted of 6 helices by 8 helices by 94 base pairs and 6 helices by 4 helices by 94 base pairs along the x and y and z directions, respectively. At the top surface of the bottom layer, we introduced four 14-nucleotide (nt) single-stranded DNA (ssDNA) handles by extending the 3' or 5' ends of four selected DNA bricks (fig. S14) (26). Extending the repeating units along the

¹Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA 02115, USA. ²Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA. ³Department of Cancer Biology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA 02115, USA. ⁴Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MA 02115, USA. ⁵IBM Almaden Research Center, San Jose, CA 95120, USA. ⁶IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598, USA. ⁷Materials Science and Engineering Division, National Institute of Standards and Technology (NIST), Gaithersburg, MD 20899, USA.

*Corresponding author. Email: py@hms.harvard.edu (P.Y.); sunwedavid1@gmail.com (W.S.) †These authors contributed equally to this work.

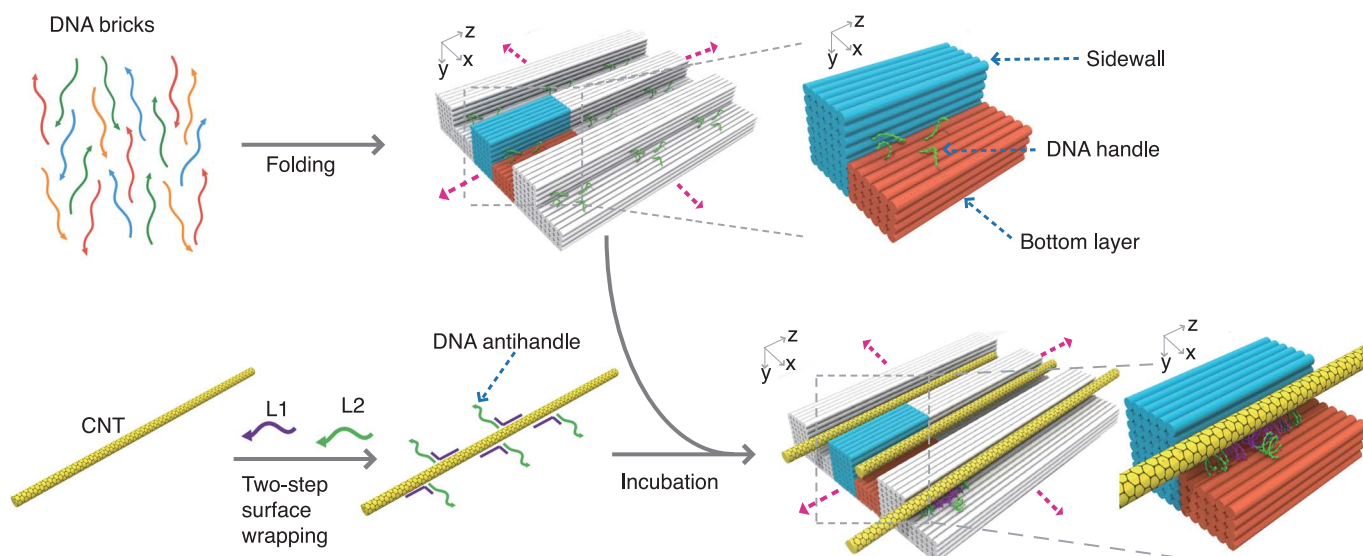


Fig. 1. Design schematic for SHINE. The blue and the orange bundles represent the sidewall and the bottom layer, respectively, within a feature-repeating unit of trench-like DNA templates. Pink arrows indicate the extension directions of the repeating units.

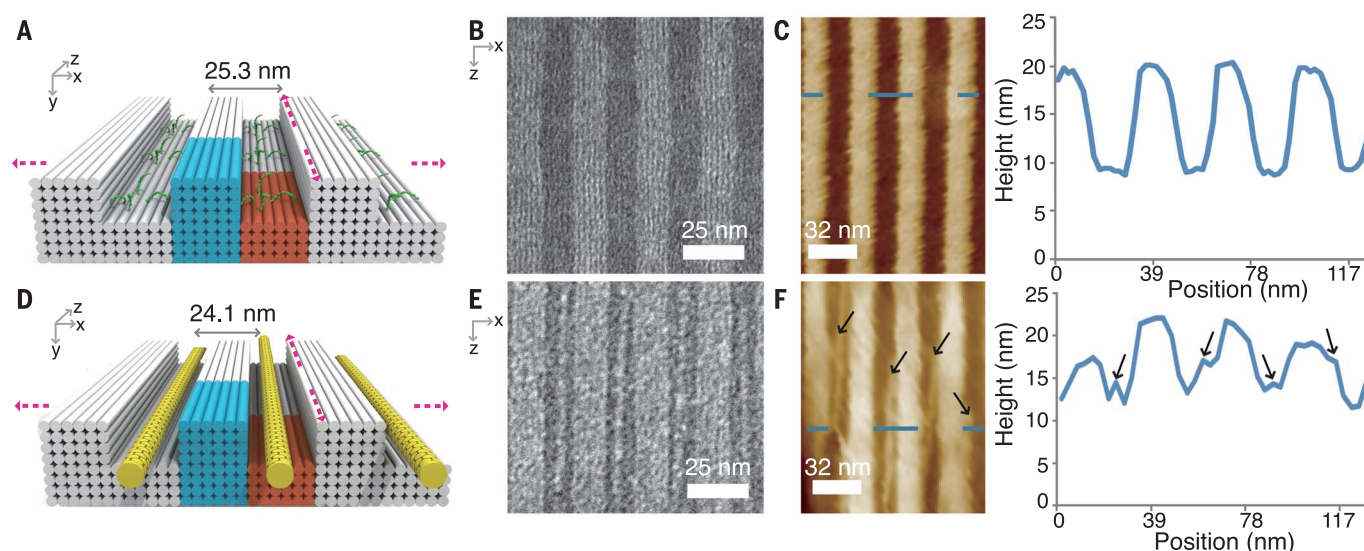


Fig. 2. Assembling CNT arrays with 24-nm inter-CNT pitch. From left to right, designs (A and D), zoomed-in TEM images along the x and z projection direction (B and E), liquid-mode AFM images along the x and z projection direction (C and F) (left), and height profiles (C and F) (right) for the DNA template (A to C) and the assembled CNT array (D to F),

respectively. Blue dashed lines [in (C) and (F), left] represent the locations for the height profile. Black arrows in the AFM image (F) indicate the assembled CNTs. See also figs. S2 to S4, S16, and S17 (26). The orientation of the assembled CNTs in (F) may be distorted by AFM tips during imaging.

x and z directions yielded DNA templates with parallel nanotrenches.

The micrometer-scale DNA templates were folded through a multistage isothermal reaction. Next, DNA antihandles were wrapped onto CNTs through noncovalent interactions (fig. S1) (27, 28). Finally, under mild conditions, the hybridization between the DNA handles and the antihandles mediated CNT assembly within the DNA nanotrenches at the prescribed inter-CNT pitch.

Transmission electron microscopy (TEM) imaging confirmed the successful formation of the designed DNA templates (Fig. 2, A and B, and figs. S2 to S4) (26), as well as the confined assembly of evenly spaced CNT arrays within the DNA nanotrenches (Fig. 2, E and F, and figs. S16 and S17). In the zoomed-out TEM images (figs. S2 and S3), the assembled DNA templates exhibited wide dimensional distributions. One typical DNA template (fig. S2C) exhibited the maximal dimensions of 1.3 μm by 200 nm in the x and z directions. In the zoomed-in TEM images, DNA templates exhibited alternative dark (bottom layer)—bright (sidewall) regions (Fig. 2B and fig. S17), and each region corresponded to six-layered DNA helices along the x direction as designed (Fig. 2A). The measured nanotrench periodicity was 25.3 ± 0.3 nm ($N = 50$ nanotrenches from 10 different templates) along the x direction after drying on the surface (corresponding to 2.1 nm diameter per dehydrated dsDNA). The ssDNA handles were not visible in the negatively stained TEM images.

After CNT assembly, we found bright parallel lines that appeared exclusively on the dark

bottom regions, indicative of the aligned CNTs along the longitudinal axis of the nanotrenches (Fig. 2, D and E, and figs. S16 and S17). The relatively larger diameter of CNTs as compared with the unwrapped CNTs was caused by the stained dsDNA layer around CNTs (fig. S15). Despite a few local twists in individual CNTs, we did not observe crossing or bundling CNT defects within the DNA nanotrenches. The measured inter-CNT pitch was 24.1 ± 1.7 nm ($N = 50$ CNTs from 10 different templates). For every two neighboring CNTs, we measured three different positions along the longitudinal axis of CNT). Slightly smaller inter-CNT pitch, compared to the x -direction periodicity of the DNA templates, was the result of statistical variance of the small sample size. The integrity of the DNA templates was not affected by CNT assembly, as indicated by the consistent six-layered DNA helices (along the x direction) in both the DNA sidewall and bottom layer (Fig. 2E).

To evaluate the pitch precision, we calculated (i) the standard deviation, (ii) the range value, (iii) the percent relative range, and (iv) the index of dispersion for count value (IDC value) for inter-CNT pitch. The range of inter-CNT pitch variation, defined as the difference between the maximum and the minimum pitch values, was 7.8 nm. The percent relative range of the inter-CNT pitch, defined as the range of inter-CNT pitch divided by the average value of inter-CNT pitch (24.1 nm), was 32%. For comparison, on a flat substrate, a range >30 nm and a percent relative range $>140\%$ have been reported for CNT arrays with similar average pitch (4).

The IDC value [defined as the standard deviation squared divided by the average pitch squared ($1/I$)] for CNT arrays (~ 40 CNTs/ μm) from SHINE was 0.005, two orders of magnitude smaller than for CNT arrays of similar density fabricated from thin-film approaches ($1/I$). Hence, by limiting the rotation of CNTs with DNA sidewalls, SHINE provided higher precision for assembling ultradense CNT arrays than flat substrate-based assembly. Similarly, SHINE produced a smaller angular deviation (less than 2° , defined as the longitudinal-axis difference between CNTs and the DNA nanotrenches) than previously obtained on flat DNA template, where $>75\%$ CNTs exhibited angular deviations $>5^\circ$ (24).

Because both DNA templates (figs. S2 and S3) and CNTs (fig. S15) exhibited uneven widths and lengths, we observed a variable number of CNTs (ranging from 4 to 15) on different templates, as well as z -direction offset for CNTs from trench to trench (fig. S17). Notably, although the width of the DNA nanotrench (12 nm) was larger than the diameter of individual CNTs, we did not observe CNT bundling within individual trenches.

We further analyzed the assembly yield of aligned CNTs by TEM counting (supplementary text S2). The assembly yield was defined as the total number of inner nanotrenches occupied by correctly assembled parallel CNT arrays divided by the total number of inner DNA nanotrenches. Partially formed DNA nanotrenches on the boundaries were excluded. A $>95\%$ assembly yield was observed for 10 randomly selected DNA templates (more than 50 inner trenches were counted, Fig. 2E

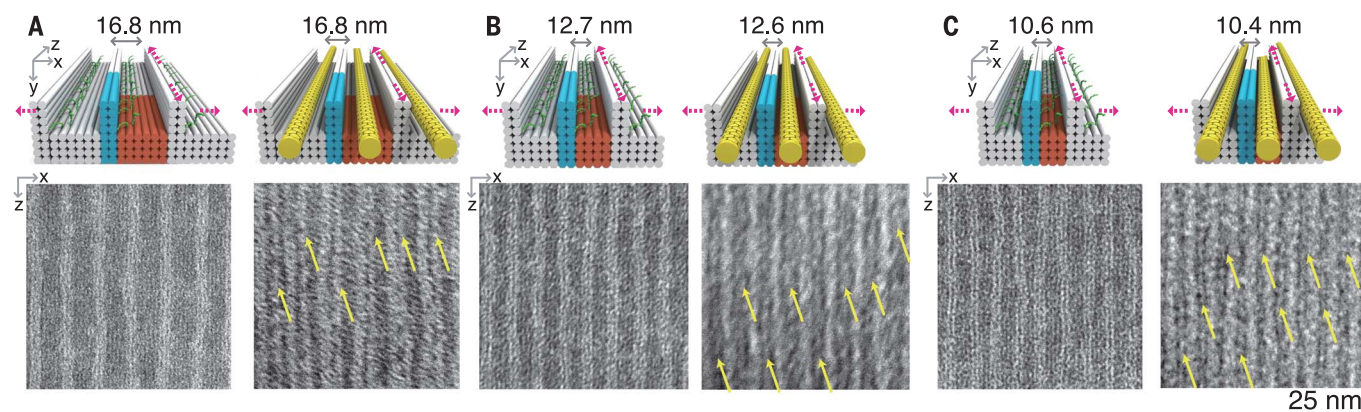


Fig. 3. Programming inter-CNT pitches with DNA brick crystal templates. (A to C) Designs (top row) and zoomed-in TEM images along the x and z projection direction (bottom row) for the DNA templates (left) and the assembled CNT arrays (right) at 16.8 nm (A), 12.6 nm (B), and 10.4 nm (C) inter-CNT pitches, respectively. Yellow arrows in the TEM images indicate the assembled CNTs. See also figs. S6 to S13 for the assembled DNA templates and figs. S19 to S24 for the assembled CNT arrays (26).

and fig. S17), and <5% of inner nanotrenches were unoccupied by CNTs (fig. S25).

In liquid-mode atomic force microscopy (AFM) images (Fig. 2F and fig. S18), we observed new peaks (with heights ~ 15 to 17 nm) within the nanotrenches (fig. S18) after CNT assembly. The height changes of the new peaks (5 to 7 nm), relative to the height of the bottom layer beneath (~ 10 nm in height, Fig. 2C), approximated the sum of dsDNA handle length (3 to 5 nm, depending on different conformations) and DNA-wrapped CNT diameter (~ 1 to 3 nm, fig. S15). Therefore, only single-layer CNTs were assembled. The ssDNA handles were not visible in the AFM images. We observed wider inter-CNT pitch (~ 32 nm) in liquid-mode AFM when compared with that from the TEM images. The pitch change was ascribed to the larger diameter of hydrated dsDNAs (2.6-nm diameter per helix) in liquid condition than of the fully dehydrated dsDNAs (2.1-nm diameter per helix under vacuum). The 32-nm inter-CNT pitch on the hydrated DNA templates could shrink to ~ 24 nm after dehydration under heat.

By programming DNA nanotrenches with different trench periodicities along the x direction, we further demonstrated prescribed scaling of inter-CNT pitches at 16.8, 12.6, and 10.4 nm (Fig. 3). Within the feature-repeating units of the small-periodicity DNA templates, we used 2 helices by 8 helices by 94 base pairs for the nanotrench sidewalls (Fig. 3, A to C, top left). In the bottom layers, 6 helices by 4 helices by 94 base pairs, 4 helices by 4 helices by 94 base pairs, and 3 helices by 4 helices by 94 base pairs were used for different nanotrench periodicities.

We assembled DNA templates and CNT arrays using approaches similar to those in Fig. 1. Assembled DNA templates exhibited measured nanotrench periodicities of 16.8 ± 0.4 nm, 12.7 ± 0.2 nm, and 10.6 ± 0.1 nm ($N = 50$

to 300 nanotrenches from 10 individual templates for each design) along x direction (Fig. 3, A to C, bottom left, and figs. S5 to S13) (26). Notably, we observed slightly twisted nanotrench sidewalls after drying in vacuum, probably because of the relatively low structural stiffness of the two-layer DNA sidewalls (29). However, the average periodicities were not affected by the twisting of the DNA sidewalls. In the zoomed-out view, different template designs showed typical dimensions of ~ 1.3 μm by 300 nm along the x and z directions (figs. S5, S6, S8, S9, S11, and S12).

After CNT assembly, parallel CNTs were aligned within the DNA nanotrenches (designs in Fig. 3, A to C, top right; TEM images in Fig. 3, A to C, bottom right; figs. S19 to S24). The inter-CNT pitches varied from 16.8 ± 1.5 nm to 12.6 ± 0.6 nm to 10.4 ± 0.4 nm, respectively ($N = 50$ to 300 CNTs from 10 individual templates for each design). Both the 10.4-nm pitch value and 0.4-nm standard deviation (smaller than the diameter of individual CNTs) were beyond current lithography-defined channel pitches (30, 31).

The IDC values were 0.008, 0.002, and 0.001, respectively—orders of magnitude smaller than those from thin-film approaches (11) (supplementary text S4.1). The range and the percent relative range of the inter-CNT pitch variation were 5.9 nm and 36%, 2.7 nm and 24%, and 1.9 nm and 18% for 16.8-, 12.6-, and 10.4-nm inter-CNT pitches, respectively. Narrower DNA nanotrenches improved the precision of CNT assembly (fig. S26). When the width of DNA nanotrenches was decreased to ~ 6 nm (in 10.4-nm pitch CNT arrays), the range value of inter-CNT pitch was decreased to <2 nm and the IDC value (0.001) improved by eightfold, compared to a 5.9-nm range value and IDC value of 0.008 in 12-nm DNA trench width (in 16.8-nm pitch CNT arrays). The angular deviations for the assembled CNTs were

less than 2° . Under the optimized buffer conditions (supplementary text S1.4), the assembly yields were over 95% (figs. S20, S22, and S24).

The synergy between electrostatic repulsions and DNA hybridization, enabled by the spatial confinement of nanotrenches, helped to eliminate the CNT assembly disorders. In the absence of DNA hybridization, CNTs could not be assembled within the DNA nanotrenches because of the electrostatic repulsions between the negatively charged CNTs and nanotrench sidewalls. The hybridization between DNA handles within the nanotrenches and the DNA antihandles wrapping around CNTs stabilized CNTs within the DNA nanotrenches and resulted in an assembly yield $>95\%$. The absence of effective DNA hybridizations in misaligned CNTs eliminated the assembly disorder by the electrostatic repulsions. The correctly assembled CNTs spatially shielded the DNA handles beneath from being accessed by other CNTs and repelled one another because of negative surface charge. Even for DNA nanotrenches (width from 6 to 12 nm) more than twofold larger than the diameter of single CNTs, we did not observe CNT bundling within individual trenches and achieved an IDC value of 0.001.

Microliter assembly solution at sub-10 pM template concentration simultaneously provided millions of assembled CNT arrays at evenly spaced pitches, demonstrating the scalability of SHINE. We further tested using thermal annealing to remove DNA templates (figs. S27 and S28) and constructed proof-of-concept transistors from parallel CNT arrays (fig. S28). The thermal decomposition of DNAs produced residual contaminations around CNTs that adversely affected the transistor performance. Thus, both low on-state current and large subthreshold swing values were recorded. By contrast, improving interface cleanliness for SHINE promotes transport performance comparable with chemical vapor

deposition-grown or polymer-wrapped CNT arrays in a follow-up study (32). Additionally, using purer semiconducting CNTs may further improve performance.

REFERENCES AND NOTES

1. G. S. Tulevski *et al.*, *ACS Nano* **8**, 8730–8745 (2014).
2. Z. Hu *et al.*, *Nat. Nanotechnol.* **11**, 559–565 (2016).
3. Q. Cao *et al.*, *Nat. Nanotechnol.* **8**, 180–186 (2013).
4. Q. Cao, S.-J. Han, G. S. Tulevski, *Nat. Commun.* **5**, 5071 (2014).
5. G. J. Brady *et al.*, *Sci. Adv.* **2**, e1601240 (2016).
6. X. He *et al.*, *Nat. Nanotechnol.* **11**, 633–638 (2016).
7. J. Wu *et al.*, *Small* **9**, 4142–4148 (2013).
8. S.-P. Han, H. T. Maune, R. D. Barish, M. Bockrath, W. A. Goddard 3rd, *Nano Lett.* **12**, 1129–1135 (2012).
9. A. Mangalum, M. Rahman, M. L. Norton, *J. Am. Chem. Soc.* **135**, 2451–2454 (2013).
10. G. J. Brady, K. R. Jinkins, M. S. Arnold, *J. Appl. Phys.* **122**, 124506 (2017).
11. M. M. Shulaker *et al.*, “High-performance carbon nanotube field-effect transistors,” in *2014 IEEE International Electron Devices Meeting (IEEE, 2014)*, pp. 33.6.1–33.6.4.
12. N. C. Seeman, *J. Theor. Biol.* **99**, 237–247 (1982).
13. N. C. Seeman, *Nature* **421**, 427–431 (2003).
14. P. W. K. Rothmund, *Nature* **440**, 297–302 (2006).
15. S. M. Douglas *et al.*, *Nature* **459**, 414–418 (2009).
16. Y. Ke, L. L. Ong, W. M. Shih, P. Yin, *Science* **338**, 1177–1183 (2012).
17. Y. Ke *et al.*, *Nat. Chem.* **6**, 994–1002 (2014).
18. S. P. Surwade *et al.*, *J. Am. Chem. Soc.* **135**, 6778–6781 (2013).
19. X. Liu *et al.*, *Nature* **559**, 593–598 (2018).
20. Z. Jin *et al.*, *Nat. Commun.* **4**, 1663 (2013).
21. A. Kuzyk *et al.*, *Nature* **483**, 311–314 (2012).
22. P. Zhan *et al.*, *Angew. Chem. Int. Ed.* **57**, 2846–2850 (2018).
23. J. B. Knudsen *et al.*, *Nat. Nanotechnol.* **10**, 892–898 (2015).
24. H. T. Maune *et al.*, *Nat. Nanotechnol.* **5**, 61–66 (2010).
25. H. Pei *et al.*, *J. Am. Chem. Soc.* **141**, 11923–11928 (2019).
26. See supplementary materials.
27. M. Zheng *et al.*, *Nat. Mater.* **2**, 338–342 (2003).
28. Z. Zhao, Y. Liu, H. Yan, *Org. Biomol. Chem.* **11**, 596–598 (2013).
29. W. Sun *et al.*, *Science* **346**, 1258361 (2014).
30. *The International Technology Roadmap for Semiconductors: 2013*; www.semiconductors.org/resources/2013-international-technology-roadmap-for-semiconductors-itsr/.
31. G. Hills *et al.*, *IEEE Trans. NanoTechnol.* **17**, 1259–1269 (2018).
32. M. Zhao *et al.*, *Science* **368**, 878–881 (2020).

ACKNOWLEDGMENTS

The authors acknowledge valuable discussions with W. L. Wilson at the Center for Nanoscale Systems of Harvard University. **Funding:** This work is supported by NSF (CMMI-1333215, CMMI-1344915,

CBET-1729397), ONR (N00014-14-1-0610, N00014-16-1-2410, N00014-18-1-2549), AFOSR (MURI FATE, FA9550-15-1-0514), and Molecular Robotics Initiative from Wyss Institute to P.Y. J.K.S., J.A.F., and M.Z. acknowledge the NIST internal fund for support. **Author contributions:** W.S. conceived, designed, and conducted the study and wrote the manuscript; J.S. designed and conducted the study and wrote the manuscript. Z. Zhao designed and conducted the study. N.A., C.R., J.T., and T.T. fabricated the FETs, characterized the transport performance, and analyzed the data. T.C. and Z. Zhou conducted the experiments and analyzed the data. J.K.S., J.A.F., and M.Z. prepared the DNA-wrapped CNTs and analyzed the data; T.S. analyzed the data and wrote the manuscript. S.-J.H. supervised FET study. W.M.S. contributed to study supervision. H.T.M. designed, conducted, and supervised FET study and wrote the paper. P.Y. conceived and supervised the study and wrote the paper. All authors reviewed, edited, and approved the manuscript. **Competing interests:** A patent based on this work was issued to W.S., J.S. and P.Y. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6493/874/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S28
References (33–37)

5 October 2019; accepted 9 April 2020
10.1126/science.aaz7440

DEVICE TECHNOLOGY

DNA-directed nanofabrication of high-performance carbon nanotube field-effect transistors

Mengyu Zhao^{1*}, Yahong Chen^{1,2*}, Kexin Wang¹, Zhaoxuan Zhang^{1,3}, Jason K. Streit⁴, Jeffrey A. Fagan⁴, Jianshi Tang⁵, Ming Zheng⁴, Chaoyong Yang², Zhi Zhu^{2†}, Wei Sun^{1†}

Biofabricated semiconductor arrays exhibit smaller channel pitches than those created using existing lithographic methods. However, the metal ions within biolattices and the submicrometer dimensions of typical biotemplates result in both poor transport performance and a lack of large-area array uniformity. Using DNA-templated parallel carbon nanotube (CNT) arrays as model systems, we developed a rinsing-after-fixing approach to improve the key transport performance metrics by more than a factor of 10 compared with those of previous biotemplated field-effect transistors. We also used spatially confined placement of assembled CNT arrays within polymethyl methacrylate cavities to demonstrate centimeter-scale alignment. At the interface of high-performance electronics and biomolecular self-assembly, such approaches may enable the production of scalable biotemplated electronics that are sensitive to local biological environments.

In projected high-performance, energy-efficient field-effect transistors (FETs) (1, 2), evenly spaced small-pitch (where pitch refers to the spacing between two adjacent channels within an individual FET) semiconductor channels are often required. Smaller channel pitch leads to higher integration density and on-state performance, but it has the risk of enhanced destructive short-range screening and electrostatic interactions in low-dimensional semiconductors, such as carbon nanotubes (CNTs) (3). Evenly spaced alignment minimizes the channel disorder that affects the switching between on and off states (4). Therefore, although high-density CNT thin films exhibit on-state performance comparable to that of Si FETs (5, 6), degraded gate modulation and increased subthreshold swing (3, 5) are observed because of the disorder in the arrays.

Biomolecules such as DNAs (7, 8) can be used to organize CNTs into prescribed arrays (9–11). On the basis of the spatially hindered integration of nanotube electronics (SHINE), biofabrication further scales the evenly spaced channel pitch beyond lithographic feasibility (12). However, none of the biotemplated CNT FETs (12–14) have exhibited performance comparable to that of those constructed with

lithography (15) or thin-film approaches (3, 5, 6, 16–18). Additionally, during the surface placement of biotemplated materials, broad orientation distributions (19) prevent their large-scale alignment.

In this work, we showed that small regions of nanometer-precise biomolecular assemblies can be integrated into the large arrays of solid-state high-performance electronics. We used the parallel semiconducting CNT arrays assembled through SHINE as model systems (12). At the FET channel interface, we observed lower on-state performance induced by high concentrations of DNA and metal ions. Using a rinsing-after-fixing approach, we eliminated the contamination without degrading CNT alignment. On the basis of the uniform inter-CNT pitch and clean channel interface, we constructed solid-state multichannel PMOS (p-channel metal-oxide semiconductor) CNT FETs that displayed high on-state performance and fast on-off switching simultaneously. Using lithography-defined polymethyl methacrylate (PMMA) cavities to spatially confine the placement of the CNT-decorated DNA templates, we demonstrated aligned arrays with prescribed geometries over a 0.35-cm²-area substrate. Building high-performance, ultrascaled devices at the biology-electronics interface may enable

diverse applications in the post-Si era, such as multiplexed biomolecular sensors (20) and three-dimensional (3D) FETs with nanometer-to-centimeter array scalability.

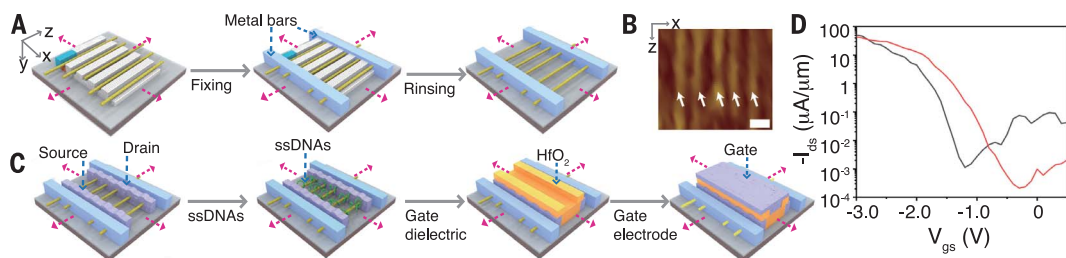
We assembled DNA-templated CNT arrays using DNA-based SHINE (12). We applied a rinsing-after-fixing approach (Fig. 1A) to remove the DNA templates. Starting from the surface-deposited DNA-templated CNT arrays, both ends of the DNA-templated CNT arrays were first fixed onto a Si wafer with deposited metal bars (first step in Fig. 1A). DNA templates and high-concentration metal salts (1 to 2 M) within the DNA helices were gently removed through sequential rinsing with water and low-concentration H₂O₂ (second step in Fig. 1A and fig. S5). The inter-CNT pitch and the alignment quality of the assembled CNTs were not degraded during the rinsing process (Fig. 1B and figs. S3 and S4) (21).

To explore the effect of single-stranded DNAs (ssDNAs) at the channel interface, we first fabricated the source and drain electrodes onto the rinsed CNT arrays (Fig. 1C, left). Next, ssDNAs were introduced exclusively into the predefined channel area (first step in Fig. 1C; channel length ~200 nm). Finally, a gate dielectric of HfO₂ and a gate electrode of Pd were sequentially fabricated (second and third steps in Fig. 1C and fig. S6).

Out of 19 FETs we constructed, 63% (12 of 19) showed typical gate modulation (on-state

Fig. 1. Multichannel CNT FETs with ssDNAs at channel interface.

(A) Design schematic for the rinsing-after-fixing approach. Pink arrows indicate the extension direction of DNA templates and the assembled CNTs. (B) Zoomed-in AFM image along the x and z projection direction for CNT arrays after template removal. White arrows indicate the assembled CNTs. Scale bar, 25 nm. See also figs. S3 and S4 (21). (C) Design schematic for introducing ssDNAs at channel interface and FET fabrication. (D) The I_{ds} - V_{gs} curves [drain-to-source current density (I_{ds}) versus V_{gs} plotted in logarithmic at a V_{ds} of -0.5 V] for a multichannel DNA-containing CNT FET before (black line) and after (red line) thermal annealing. See also fig. S7.



¹Key Laboratory for the Physics and Chemistry of Nanodevices and Center for Carbon-Based Electronics, Department of Electronics, Peking University, Beijing 100871, China. ²Collaborative Innovation Center of Chemistry for Energy Materials, The MOE Key Laboratory of Spectrochemical Analysis and Instrumentation, State Key Laboratory of Physical Chemistry of Solid Surfaces, Department of Chemical Biology, College of Chemistry and Chemical Engineering, Xiamen University, Xiamen 361005, China. ³State Key Laboratory of Fine Chemicals, Dalian University of Technology, Dalian 116024, China. ⁴Materials Science and Engineering Division, National Institute of Standards and Technology (NIST), Gaithersburg, MD 20899, USA. ⁵Institute of Microelectronics, Beijing Innovation Center for Future Chips (ICFC), Beijing National Research Center for Information Science and Technology (BNRist), Tsinghua University, Beijing 100084, China.

*These authors contributed equally to this work.

†Corresponding author. Email: sunw@pku.edu.cn (W.S.); zhuzhi@xmu.edu.cn (Z.Z.)

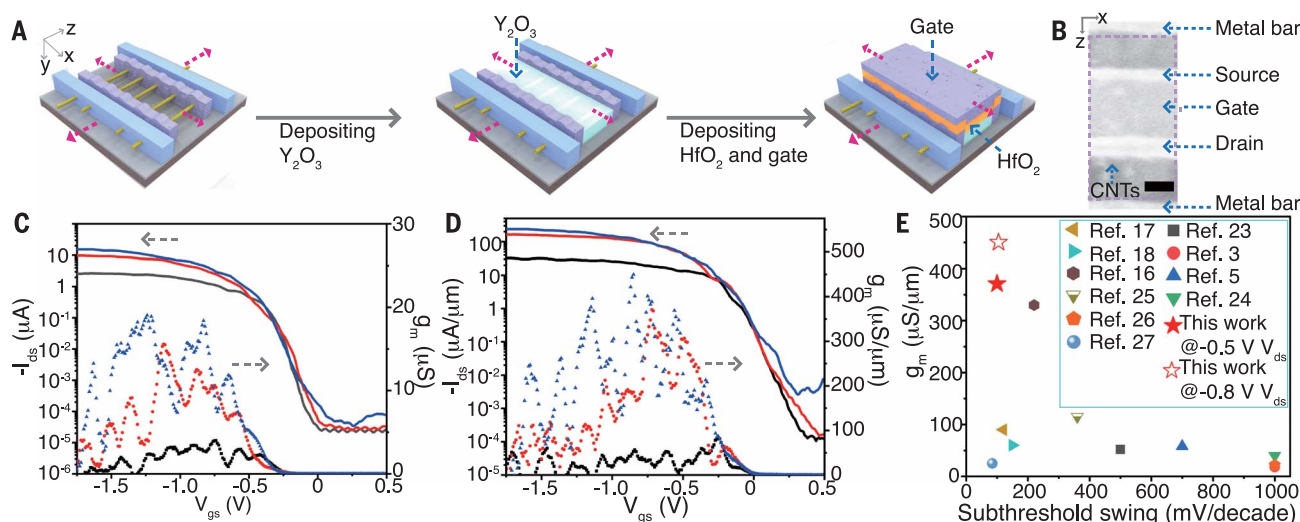


Fig. 2. Constructing top-gated high-performance CNT FETs. (A) Design schematic for the fabrication of top-gated DNA-free FETs. (B) Zoomed-in SEM image along the x and z projection direction for the constructed multichannel CNT FET. Scale bar, 100 nm. See also fig. S11 (21). (C) and (D) The I_{ds} - V_{gs} curves (solid lines, plotted in logarithmic scale corresponding to left axis) and g_m - V_{gs} curves (dotted lines, plotted in linear scale corresponding to right axis) for single-channel (C) and

multichannel (D) CNT FETs. Blue, red, and black colors in (C) and (D) represent a V_{ds} of -0.8 , -0.5 , and -0.1 V, respectively. Gray arrows indicate the corresponding axes. See also figs. S9 and S12. (E) Benchmarking of the current multichannel CNT FET in (D) with other reports of high-performance CNT FETs. Device performances from previous publications (3, 5, 16–18, 23–27) are obtained at a V_{ds} of -0.5 V and channel lengths ranging from 100 to 500 nm. See also figs. S17 and S18.

current density divided by off-state current density, I_{on}/I_{off} , exceeded 10^3 ; fig. S7). The other seven devices exhibited $I_{on}/I_{off} < 5$, which was caused by the presence of metallic CNTs within the array. At a drain-to-source bias (V_{ds}) of -0.5 V, one typical multichannel DNA-containing CNT FET (Fig. 1D) exhibited a threshold voltage (V_{th}) of ~ -2 V, an I_{on} of $50 \mu A/\mu m$ (normalized to the inter-CNT pitch) at a gate-to-source bias (V_{gs}) of -3 V, a subthreshold swing of 146 mV per decade, a peak transconductance (g_m) of $23 \mu S/\mu m$, and an on-state conductance (G_{on}) of 0.10 mS/ μm . Statistics over all of the 12 operational FETs exhibited a V_{th} distribution of -2 ± 0.10 V, an I_{on} of 4 to $50 \mu A/\mu m$, and a subthreshold swing of 164 ± 44 mV per decade (fig. S7A). The transport performance was stable during repeated measurements (fig. S7C).

We annealed the above DNA-containing FETs at $400^\circ C$ for 30 min under vacuum to thermally decompose ssDNAs (22), and we then recharacterized the transport performance. Compared with the unannealed samples, thermal annealing (Fig. 1D and figs. S7 and S16) slightly shifted the average V_{th} (~ -0.35 V, for a V_{th} of -1.65 ± 0.17 V after annealing) and increased the average subthreshold swing by ~ 70 mV per decade (subthreshold swing of 230 ± 112 mV per decade after annealing). Other on-state performance metrics, including g_m and G_{on} , as well as FET morphology, did not substantially change after annealing.

To build high-performance CNT FETs from biotemplates, we deposited a composite gate

dielectric (Y_2O_3 and HfO_2) into the rinsed channel area instead of introducing ssDNAs (Fig. 2, A and B, and figs. S10 and S11) (21). Of all the FETs constructed, 54% (6 of 11) showed gate modulation (fig. S12). The other 5 of 11 FETs contained at least one metallic CNT within the channel (fig. S15). Using an identical fabrication process, we also constructed another nine operational single-channel DNA-free CNT FETs for comparing transport performance (fig. S8). The single-channel CNT FET (channel length ~ 200 nm) with the highest on-state performance exhibited an on-state current of $10 \mu A$ per CNT (V_{ds} of -0.5 V) at the thermionic limit of subthreshold swing (i.e., 60 mV per decade; Fig. 2C and fig. S9).

At a V_{ds} of -0.5 V, the multichannel DNA-free CNT FET (channel length ~ 200 nm, inter-CNT pitch of 24 nm) with the highest on-state performance (Fig. 2D and fig. S13) exhibited a V_{th} of -0.26 V, an I_{on} of $154 \mu A/\mu m$ (at a V_{gs} of -1.5 V), and a subthreshold swing of 100 mV per decade. The g_m and G_{on} values were 0.37 and 0.31 mS/ μm , respectively. The noise in the g_m - V_{gs} curves may originate from thermal noise, or disorder and scattering within the composite gate construct. The on-state current further increased to $\sim 250 \mu A/\mu m$, alongside a g_m of 0.45 mS/ μm and a subthreshold swing of 110 mV per decade, at a V_{ds} of -0.8 V.

At a similar channel length and V_{ds} (-0.5 V), we benchmarked the transport performance (g_m and subthreshold swing) against that of conventional thin-film FETs using chemical vapor deposition (CVD)-grown or polymer-

wrapped CNTs (3, 5, 16–18, 23–27) (Fig. 2E and figs. S17 and S18). Both high on-state performance (a g_m of ~ 0.37 mS/ μm) and fast on-off switching (a subthreshold swing of ~ 100 mV per decade) could be simultaneously achieved within the same solid-state FET, whereas thin-film CNT FETs with a similar subthreshold swing (~ 100 mV per decade) exhibited a $>50\%$ smaller g_m .

When the channel length was scaled to 100 nm, we achieved an I_{on} of $300 \mu A/\mu m$ (at a V_{ds} of -0.5 V and a V_{gs} of -1.5 V) and a subthreshold swing of 160 mV per decade (fig. S14). Both the G_{on} and the g_m values were thus promoted to 0.6 mS/ μm . The DNA-free CNT FETs exhibited comparable I_{on} to that of thin-film FETs from aligned CVD-grown CNT arrays (28, 29), even at 60% smaller CNT density [~ 40 CNTs/ μm versus >100 CNTs/ μm in (28, 29)]. The effective removal of the contaminations, such as DNA and metal ions, and the shorter channel length contributed to the high I_{on} . Notably, a previous study had fixed CNTs directly with the source and drain electrodes (13). Because contamination could not be fully removed from the electrode contact areas, the on-state performance (g_m and G_{on}) decreased by a factor of 10.

Furthermore, the subthreshold swing difference between the multichannel (average value of 103 mV per decade) and the single-channel CNT FETs (average value of 86 mV per decade in fig. S9) was reduced to 17 mV per decade. Theoretical simulations suggest that, under identical gate constructs, the uneven diameter of CNTs (6) and the alignment disorder

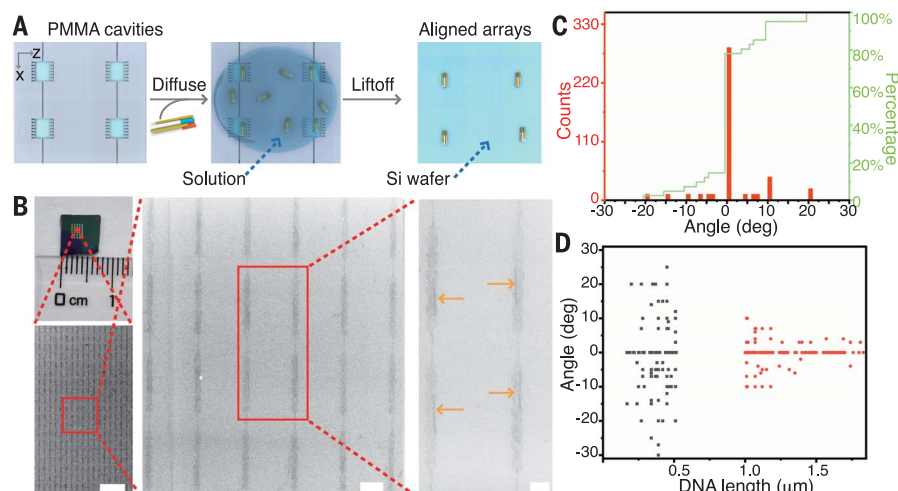


Fig. 3. Centimeter-scale oriented placement of fixed-width arrays. (A) Design schematic for the oriented placement of the fixed-width CNT-decorated DNA templates on a Si substrate. From left to right, the panels show fabricating cavities on a spin-coated PMMA layer, depositing CNT-decorated DNA templates onto the PMMA cavities, and liftoff to remove the PMMA layer. (B) From left to right, zoomed-out and zoomed-in optical and SEM images of the aligned structures on the Si wafer after PMMA liftoff. The scale bars in the bottom left, middle, and right images are 10, 1, and 0.5 μm , respectively. The red rectangles indicate the selected areas for zoomed-in views. The yellow arrows in the right panel indicate the aligned arrays. See also fig. S21 (21). (C) The statistics of counts (left, red axis) and the cumulative percentages (right, green axis) for the aligned structures in (B) at each specific orientation. (D) Plot of angular distributions of the aligned arrays versus the lengths of the DNA templates.

(including crossing CNTs) (5) raise the subthreshold swing (4). We observed a wide diameter distribution of the DNA-wrapped CNTs in atomic force microscopy (AFM) images (fig. S2) and transmission electron microscopy images (fig. S1). Hence, the small subthreshold swing difference above indicated that effective gate modulation and evenly spaced CNT alignment were achieved using SHINE (12) (i.e., the absence of crossing or bundling CNTs within the channel area).

Statistics across all the operational multichannel DNA-free FETs exhibited a V_{th} of -0.32 ± 0.27 V, an I_{on} of 25 to 154 $\mu\text{A}/\mu\text{m}$ (at a V_{ds} of -0.5 V and a V_{gs} of -1.5 V), and a subthreshold swing of 103 ± 30 mV per decade. Different amounts of narrow CNTs (i.e., those with diameters <1 nm) within FETs led to the wide distribution of I_{on} . Because the Schottky barrier and the bandgap increase with narrower CNT diameters, lower CNT conductance is often observed in narrow CNTs than in those with diameters >1.4 nm (30, 31).

When comparing the transport performance differences between DNA-containing and DNA-free FETs (fig. S16), we observed a largely negatively shifted V_{th} (-2 versus -0.32 V), a higher drain-to-source current density (I_{ds}) at a positive V_{gs} (mostly 10 to 200 versus 0.1 to 10 $\text{nA}/\mu\text{m}$), and a more than one order of magnitude smaller g_{m} (4 to 50 versus 70 to 370 $\mu\text{S}/\mu\text{m}$). Thus, high-concentration ssDNAs and metal ions within

multichannel FETs deteriorated the transport performance. Thermal annealing did not fully eliminate the adverse effect because of the presence of insoluble and nonsublimable annealing products, such as metal phosphates (22).

When CNT-decorated DNA templates were deposited onto a flat Si wafer, random orientations of DNA templates were formed through unconfined surface rotation. We solved this issue by using 3D polymeric cavities to confine the surface orientation during large-area placement. We first assembled fixed-width CNT arrays (fig. S19) (21) with a prescribed inter-CNT pitch of 16 nm (two CNTs per array). Next, in a typical 500 μm -by-500 μm write-field on the PMMA-coated Si substrate (with >20 write-fields on a 0.35- cm^2 substrate), we fabricated densely aligned crenellated parapet-like PMMA cavities (cavity density of $\sim 2 \times 10^7$ cavities/ cm^2 ; fig. S20). The minimum and the maximum designed widths of an individual cavity along the z direction were 180 and 250 nm, respectively.

After DNA deposition and PMMA liftoff (Fig. 3A), $>85\%$ of the initial cavities (~ 600 cavities were counted) were occupied by DNA templates (Fig. 3B and fig. S21). The measured angular distribution—defined as the difference between the longitudinal axis of the DNA templates and the x direction of the substrate—was 56% within $\pm 1^\circ$ and 90% within $\pm 7^\circ$ (Fig. 3C), per scanning electron microscopy (SEM)-based counting of all of the remaining DNA

templates within the 600 cavity sites. This value included improvable effects from the fabrication defects of PMMA cavity sites, the variation during DNA placement, and any disturbance from PMMA liftoff. Notably, the angular distribution was still improved compared with previous large-scale placement of DNA-templated materials (19). CNTs were not visible under SEM because they were embedded within the DNA trenches and shielded from the SEM detector by DNA helices.

Both the lengths of the DNA templates and the aspect ratio of the PMMA cavities affected the angular distribution. Longer DNA templates (with lengths >1 μm) exhibited narrower angular distribution ($0^\circ \pm 3.4^\circ$ in Fig. 3D) than those of shorter DNA templates (with lengths <500 nm, $1^\circ \pm 11^\circ$ in Fig. 3D). Additionally, PMMA cavities with a higher length-to-width aspect ratio (i.e., 10 in Fig. 3B and fig. S20) provided better orientation controllability than those with a lower aspect ratio (i.e., 1 in fig. S22). Hence, longer DNA templates, as well as a higher length-to-width aspect ratio of PMMA cavities, were beneficial in improving the angular distribution. Because PMMA cavities were wider than the DNA templates, we observed up to three DNA templates, as well as the offset of DNA templates along the x and z directions, within a few PMMA cavities. Notably, DNA templates did not fully cover the individual PMMA cavities, even for a saturated DNA solution.

Two-dimensional hydrophilic surface patterns, with shape and dimensions identical to those of the DNA structures, could direct the orientation of the deposited DNA structures (32). However, it is difficult to design patterns adaptive to DNA templates with variable lengths. In contrast, effective spatial confinement relies mainly on the lengths of the DNA templates and the aspect ratio of PMMA cavities and is applicable to irregular template lengths. Therefore, the anisotropic biotemplated CNT arrays with uneven lengths could be aligned along the longitudinal direction of the cavities (supplementary text section S4.1 and fig. S23) (21).

To further promote the on-state performance, scaling the inter-CNT pitch into <10 nm may be beneficial. However, at 2-nm inter-CNT pitch, the enhanced electrostatic interactions may affect the on-off switching. Therefore, the correlation between the inter-CNT pitch and performance metrics of CNT FETs needs to be verified. Combined with large-area fabrications through conventional lithography and directed assembly of block copolymers, biomolecular assembly could provide a high-resolution paradigm for programmable electronics over large areas. The hybrid electronic-biological devices may also integrate electrical stimuli and biological inputs and outputs, producing ultrascaled sensors or bioactuators.

REFERENCES AND NOTES

1. "2013 International Technology Roadmap for Semiconductors (ITRS)" (Semiconductor Industry Association, 2013); www.semiconductors.org/resources/2013-international-technology-roadmap-for-semiconductors-itsr/.
2. G. Hills *et al.*, *IEEE Trans. Nanotechnol.* **17**, 1259–1269 (2018).
3. Q. Cao *et al.*, *Nat. Nanotechnol.* **8**, 180–186 (2013).
4. G. J. Brady, K. R. Jinkins, M. S. Arnold, *J. Appl. Phys.* **122**, 124506 (2017).
5. G. J. Brady *et al.*, *Sci. Adv.* **2**, e1601240 (2016).
6. Q. Cao, J. Tersoff, D. B. Farmer, Y. Zhu, S.-J. Han, *Science* **356**, 1369–1372 (2017).
7. H. A. Becerril, A. T. Woolley, *Chem. Soc. Rev.* **38**, 329–337 (2009).
8. M. R. Jones, N. C. Seeman, C. A. Mirkin, *Science* **347**, 1260901 (2015).
9. S.-P. Han, H. T. Maune, R. D. Barish, M. Bockrath, W. A. Goddard 3rd, *Nano Lett.* **12**, 1129–1135 (2012).
10. A. Mangalum, M. Rahman, M. L. Norton, *J. Am. Chem. Soc.* **135**, 2451–2454 (2013).
11. H. Pei *et al.*, *J. Am. Chem. Soc.* **141**, 11923–11928 (2019).
12. W. Sun *et al.*, *Science* **368**, 874–877 (2020).
13. H. T. Maune *et al.*, *Nat. Nanotechnol.* **5**, 61–66 (2010).
14. K. Keren, R. S. Berman, E. Buchstab, U. Sivan, E. Braun, *Science* **302**, 1380–1382 (2003).
15. S.-J. Han *et al.*, *Nat. Nanotechnol.* **12**, 861–865 (2017).
16. D. Zhong *et al.*, *Nat. Electron.* **1**, 40–45 (2018).
17. C. Zhao *et al.*, *Adv. Funct. Mater.* **29**, 1808574 (2019).
18. L. Liu *et al.*, *ACS Nano* **13**, 2526–2535 (2019).
19. A. M. Hung *et al.*, *Nat. Nanotechnol.* **5**, 121–126 (2010).
20. N. Nakatsuka *et al.*, *Science* **362**, 319–324 (2018).
21. See supplementary materials.
22. F. Zhou *et al.*, *ACS Nano* **10**, 3069–3077 (2016).
23. Q. Cao, S.-J. Han, G. S. Tulevski, *Nat. Commun.* **5**, 5071 (2014).
24. J. Wu *et al.*, *Small* **9**, 4142–4148 (2013).
25. Y. Cao *et al.*, *ACS Nano* **10**, 6782–6790 (2016).
26. J. Wu, A. Antaris, M. Gong, H. Dai, *Adv. Mater.* **26**, 6151–6156 (2014).
27. D. Lee *et al.*, *ACS Nano* **10**, 10894–10900 (2016).
28. M. M. Shulaker *et al.*, in *2014 IEEE International Electron Devices Meeting* (IEEE, 2014), pp. 33.6.1–33.6.4.
29. Y. Hu *et al.*, *Nat. Commun.* **6**, 6099 (2015).
30. L. Zhang *et al.*, *J. Am. Chem. Soc.* **130**, 2686–2691 (2008).
31. L. Zhang *et al.*, *J. Am. Chem. Soc.* **131**, 2454–2455 (2009).
32. A. Gopinath, P. W. K. Rothmund, *ACS Nano* **8**, 12030–12040 (2014).

ACKNOWLEDGMENTS

Funding: W.S., M.Zha., Y.C., K.W., and Z.Zha. acknowledge the National Science Foundation of China (grant nos. 21875003, 21991134, and 61621061) and Peking University for financial support. Y.C., C.Y., and Z.Zhu. acknowledge the National Science Foundation of China (grant nos. 21775128, 21435004, and

21974113) for financial support. J.K.S., J.A.F., and M.Zhe. acknowledge the NIST internal fund. **Author contributions:** M.Zha. conducted the experiments on CNT assembly and CNT FETs and analyzed the data; Y.C. conducted the experiments on CNT assembly and centimeter-scale placement and analyzed the data; K.W. and Z.Zha. conducted the experiments on CNT assembly and analyzed the data; J.K.S., J.A.F., and M.Zhe. prepared the DNA-wrapped CNTs and analyzed the data; J.T. analyzed the data; Z.Zhu. designed and supervised the study and interpreted the data; W.S. conceived, designed, and supervised the study and interpreted the data; and all authors wrote the manuscript. **Competing interests:** Two provisional-stage patent applications were submitted by W.S. and M.Zha. (regarding FET construction) and W.S. and Y.C. (regarding large-area alignment). **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6493/878/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S23
References (33–37)

5 October 2019; accepted 9 April 2020
10.1126/science.aaz7435

POLLINATORS

Bumble bees damage plant leaves and accelerate flower production when pollen is scarce

Foteini G. Pashalidou^{1,2*}, Harriet Lambert^{1*}, Thomas Peybernes¹,
Mark C. Mescher^{1†}, Consuelo M. De Moraes^{1†}

Maintaining phenological synchrony with flowers is a key ecological challenge for pollinators that may be exacerbated by ongoing environmental change. Here, we show that bumble bee workers facing pollen scarcity damage leaves of flowerless plants and thereby accelerate flower production. Laboratory studies revealed that leaf-damaging behavior is strongly influenced by pollen availability and that bee-damaged plants flower significantly earlier than undamaged or mechanically damaged controls. Subsequent outdoor experiments showed that the intensity of damage inflicted varies with local flower availability; furthermore, workers from wild colonies of two additional bumble bee species were also observed to damage plant leaves. These findings elucidate a feature of bumble bee worker behavior that can influence the local availability of floral resources.

Ongoing environmental change is exacerbating phenological mismatches between plants and pollinators (1–9), with the greatest detrimental effects on pollinator populations occurring in early spring (10, 11). Such mismatches may be particularly challenging for bumble bees, which establish new colonies each spring (7, 12–14). One recent study found that the timely availability of floral resources within 1 km of founding colonies strongly influenced colony survival and performance (9), while another showed that resource shortfalls during a critical period of early colony development have profound effects on later reproductive success (14). The fitness implications of phenological mismatches presumably favor adaptations that ensure synchrony (7, 13–20), and previous research on social bees has documented changes in life history and social organization associated with variation in the seasonal availability of floral resources (13, 15–22). This study explores whether bumble bee workers may themselves influence local flower availability by damaging the leaves of flowerless plants and thereby accelerating flower production.

Initial behavioral observations with four plant species (*Brassica oleracea*, *B. nigra*, *Solanum elaeagnifolium*, and *S. melongena*) revealed that bumble bee (*Bombus terrestris*) workers use their proboscises and mandibles to cut distinctively shaped holes in plant leaves, with each damage event taking only a few seconds (Fig. 1 and movie S1). However, we saw no clear evidence that bees were actively feeding on leaves or transporting leaf material back to the hive. We therefore hypothesized that dam-

age inflicted on plant leaves might influence subsequent flower production. Although stress is known to influence flowering time in many plant species (23), most previous work has focused on abiotic stressors, and few studies have explored effects of leaf damage inflicted by insects (23, 24).

To determine whether damage inflicted by *B. terrestris* workers influences flowering, we conducted experiments with two distantly related plants, *S. lycopersicum* and *B. nigra*, and compared the flowering time of bee-damaged plants to that of undamaged plants and plants subjected to similar amounts of damage inflicted mechanically. Plants given bee-damage treatments were individually placed in mesh cages with pollen-deprived *B. terrestris* colonies and removed when worker bees had made 5 to 10 leaf holes. Each damaged plant was then paired with a plant given the mechanical-damage treatment, in which we replicated the pattern of bee-inflicted damage as closely as possible using metal forceps and a razor (see supplementary materials for further methodological details). These experiments revealed highly significant treatment effects on flowering time in both species, with bee-damaged plants flowering earlier than those that underwent the other treatments (Fig. 2). Furthermore, the acceleration of flowering elicited by bee-inflicted damage was substantial: In *S. lycopersicum*, the average flowering time of bee-damaged plants was 30 days earlier than that of undamaged plants and 25 days earlier than that of mechanically damaged plants. In *B. nigra*, the same comparisons yielded differences of 16 and 8 days, respectively (Fig. 2). Pairwise treatment comparisons [via a generalized linear model (GLM)] confirmed highly significant effects of bee damage compared with other treatments for both plant species (fig. S1). The mean flowering time of mechanically damaged plants was intermediate between bee-damaged and undamaged

treatments in both species but differed significantly from that of undamaged plants only for *B. nigra* (fig. S1).

Initial observations of laboratory colonies suggested that damaging behavior was influenced by the availability of pollen, the only food given to bumble bee larvae and an important protein source for workers (25). To test this hypothesis, we conducted a laboratory assay in which *B. terrestris* microcolonies were either given abundant pollen resources (within the hive) or deprived of pollen. After subjecting colonies to these treatments for several days (Fig. 3A), we presented them with flowerless *B. nigra* plants within mesh enclosures then replaced plants daily and quantified the proportion of damaged leaves. Trials with paired pollen-satiated and pollen-deprived colonies were conducted in parallel, and the treatment for each colony was then reversed (Fig. 3A). The results of this experiment revealed that rates of damaging behavior were markedly higher under pollen limitation regardless of the order in which treatments were presented (Fig. 3B). These findings are reinforced by further observations showing that workers from pollen-satiated colonies consistently inflict only minor amounts of leaf damage (table S2), whereas higher levels of damaging behavior can consistently be elicited by depriving microcolonies of pollen for ~3 days, as in our flowering-time experiments.

The laboratory studies described above demonstrate that the leaf-damaging behavior of *B. terrestris* workers accelerates flowering and is itself influenced by the availability of floral resources. However, these results do not exclude the possibility that the observed behavior might be an artifact of our experimental conditions, in which bees were confined in cages that prevented them from going farther afield to forage for floral resources. To address this possibility, we next conducted a series of seminatural outdoor experiments by establishing experimental *B. terrestris* colonies and plant patches on rooftops at the Zentrum campus of ETH Zürich (Zürich, Switzerland) in 2018 and 2019.

The primary goal of our 2018 study was to determine whether bees would damage flowerless plants near the hive when they had the option of foraging farther afield. A secondary goal was to test the prediction that the frequency of damaging behavior would decline over time as floral resources became more abundant in the surrounding environment. This study was carried out in two phases, the first beginning on 26 March 2018, slightly earlier than the time at which wild bumble bee queens would likely be terminating diapause (19), and continuing to 25 May. Phase 1 thus extended over a period during which surrounding flower resources increased from initially low levels (26, 27) (fig. S2). During this phase, we positioned a *B. terrestris* microcolony

¹Department of Environmental Systems Sciences, ETH Zürich, 8092 Zürich, Switzerland. ²UMR Agronomie, INRA, AgroParisTech, Université Paris-Saclay, 78850 Thiverval-Grignon, France.

*These authors contributed equally to this work.

†Corresponding author. Email: consuelo.demoraes@usys.ethz.ch (C.M.D.M.); mescher@usys.ethz.ch (M.C.M.)

adjacent to a patch of 36 flowerless plants (from six different species; see supplementary materials for details). The colony and plants were replaced every 3 weeks, and we assessed plant damage and bee activity on weekdays (except when bees were inactive

owing to weather). Bumble bee workers damaged plants in our experimental patches throughout this phase of the experiment (Fig. 4A); furthermore, as predicted, we observed a strong decline in damage by the end of April, at which point local floral resources

were relatively abundant (26, 27) (fig. S2). These results thus complement the findings from our laboratory studies showing that damaging behavior is influenced by pollen availability. As we only placed one microcolony at a time during this study, some temporal variation in damage might alternatively be explained by colony effects, although our laboratory experiments revealed minimal intercolony variation in damaging behavior (Fig. 3B and table S2).

The second phase of our 2018 outdoor experiment started on 4 June 2018 and continued through 20 July. This phase replicated the design of the first, except that we placed 100 plants in flower adjacent to the focal patch of flowerless plants (~1 m from the hive) to ensure the availability of local floral resources (see supplementary materials for details). The intensity of damage inflicted during this phase was far lower than that observed during phase 1 (Fig. 4A), which again is consistent with the prediction that abundant floral resources reduce damaging behavior. Furthermore, in addition to *B. terrestris*, workers from at least two other bumble bee species damaged flowerless plants in our experimental patch during phase 2. Specifically, four *B. lapidarius* and three *B. lucorum* workers were observed inflicting leaf damage in a manner similar to *B. terrestris* (table S3). These observations also confirm that damaging behavior is not exclusive to commercial bumble bee hives but

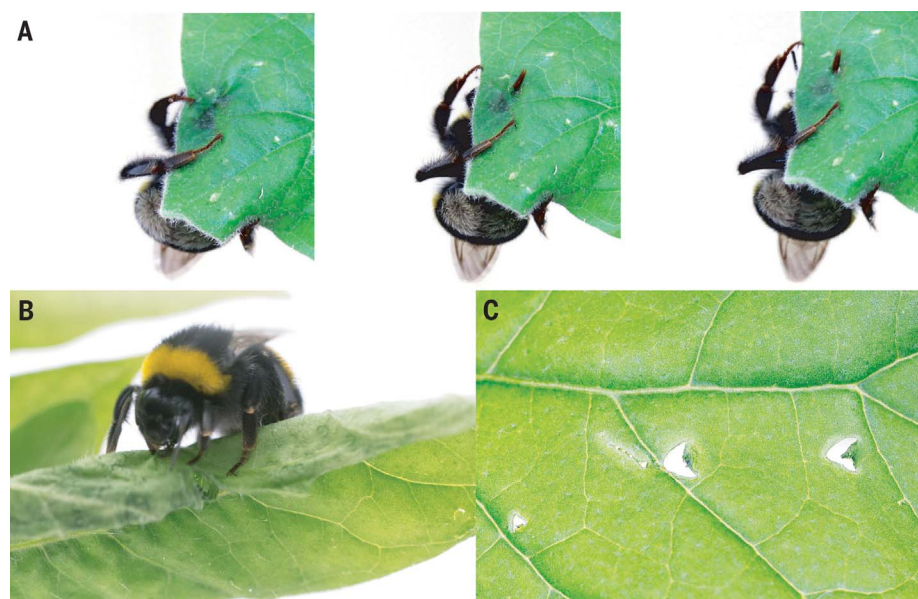


Fig. 1. *Bombus terrestris* workers damaging *Solanum melongena* leaves. (A) Sequential images of a worker penetrating a leaf with its proboscis (taken over ~3 s). (B) A worker cutting into a leaf with its mandibles. (C) Characteristic bee-inflicted damage.

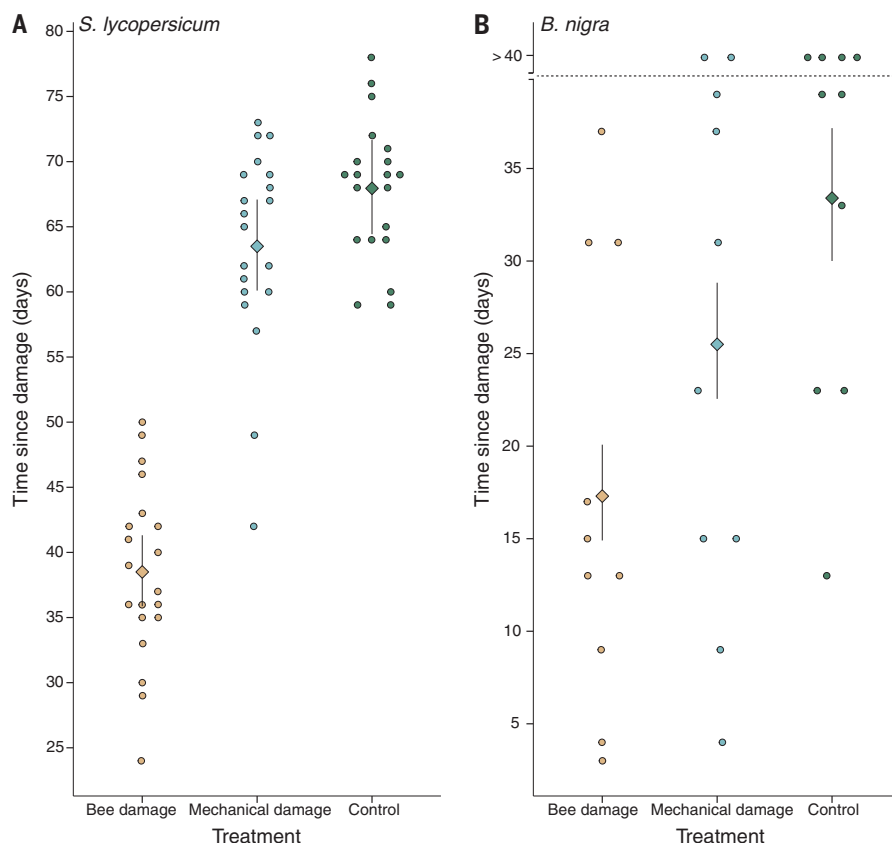


Fig. 2. Effects of bee damage on flowering time.

Flowering time of *Solanum lycopersicum* and *Brassica nigra* plants subjected to three treatments: (i) bee damage, (ii) mechanical damage, and (iii) no damage (control). Bars represent least squares means \pm confidence intervals. Data were converted to a binary response variable (flowering or not flowering, for each plant on each day), and treatment effects were analyzed using generalized additive models (GAMs), avoiding assumptions of linearity. Interactions between “time since damage” and plant were modeled by smooth functions (table S1). (A) *S. lycopersicum*: All treatments had highly significant effects on flowering time (bee damage: $P < 0.001$, estimate = 13.697; mechanical damage: $P < 0.001$, estimate = 9.131; control: $P < 0.001$, estimate = -24.279). The interaction between time and plant (smooth term) was also significant: $P < 0.001$, estimated degrees of freedom (edf) = 0.974, Chi square (Chi.sq) = 39.49. GAM explained 77% of deviance, with coefficient of determination (R^2) = 0.794 over 4800 observations. $n = 20$ plants per treatment. (B) *B. nigra*: All treatments had highly significant effects on flowering time (bee damage: $P < 0.001$, estimate = 2.955; mechanical damage: $P < 0.001$, estimate = 2.601; control: $P < 0.001$, estimate = -5.747). The interaction between time and plant (smooth term) was also significant: $P < 0.001$, edf = 0.994, Chi.sq = 181.9. GAM explained 48% of the deviance, with R^2 = 0.532 over 1200 observations. $n = 10$ plants per treatment.

is also exhibited by wild bees. Indeed, some *B. terrestris* workers observed damaging plants may also have come from wild colonies, although these could not be reliably distinguished from bees from our experimental

colonies. It is also notable that only bumble bees were observed to land on flowerless plants within our experimental patches, although other bees, including honey bees (*Apis mellifera*) and common furry bees (*Lasioglossum calceatum*),

were frequent visitors to the nearby patch of plants in flower (fig. S3).
In 2019, we conducted a second outdoor study that differed from the 2018 study in several respects. This second study used young

Fig. 3. Leaf damage by pollen-satiated and pollen-deprived bumble bees. (A) Experimental setup:

Two colonies were used in parallel for each trial, one of which was initially “pollen satiated” (PS) and then switched to “pollen deprived” (PD), while the other experienced the same treatments in the reverse order. Asterisks indicate days (1 and 8) when hives were weighed and new diet treatments were implemented. Adjustment (“A”) periods (days 2 to 4 and days 9 to 11) allowed colonies to acclimate to treatments. (B) Bees from pollen-deprived colonies inflicted higher levels of damage on plant leaves [*F* value (*F*) = 258, *df* = 32, *R*² = 0.8801, *P* < 0.001, GLM], regardless of the order of the treatment. Additionally, the observed effects did not differ significantly among colonies (*F* = 258, *df* = 32, *P* = 0.732, GLM). Boxes represent the first to third quartile of the interquartile range, horizontal lines within boxes are medians, and the whiskers are the minimum and maximum values. Six hives, three replicates, and 72 plants. ****P* ≤ 0.001; ns, not significant.

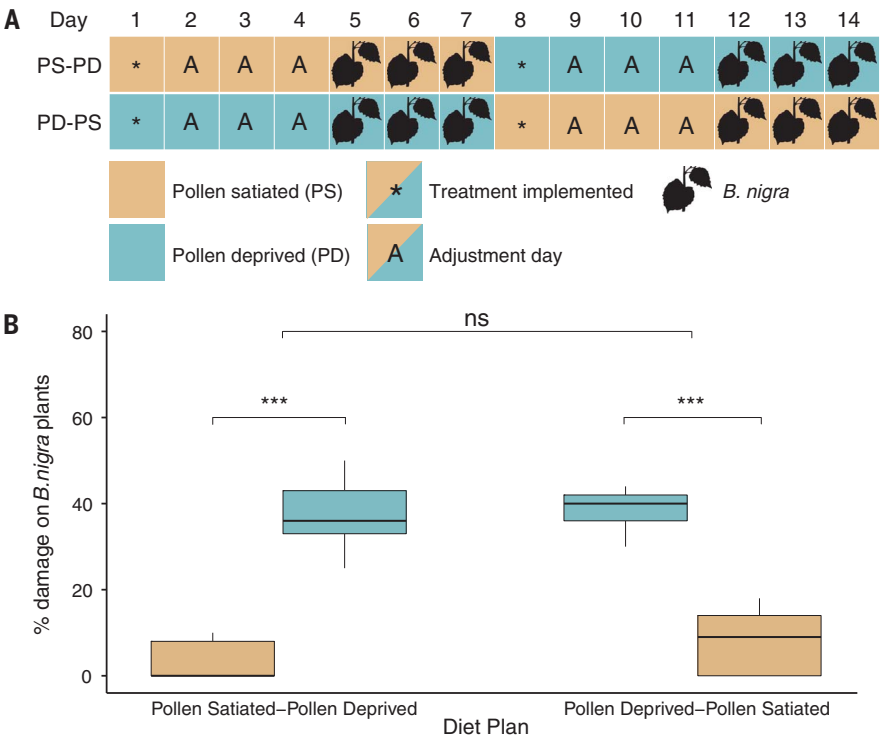
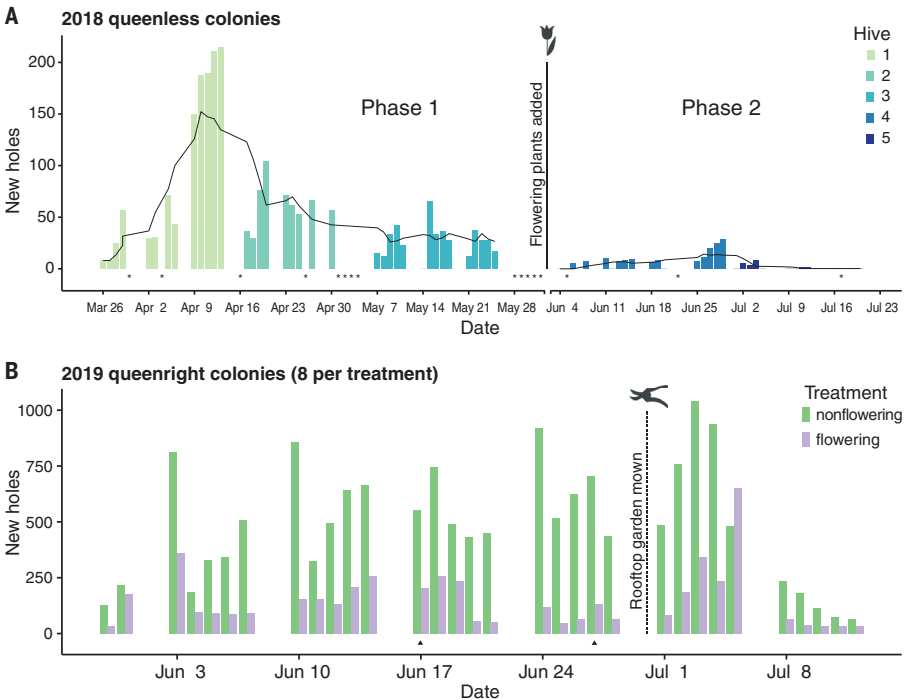


Fig. 4. Daily measurements of damage (total number of new leaf holes) by *B. terrestris* workers in two outdoor experiments. (A) 2018 experiment

with queenless microcolonies. In phase 1, only flowerless plants were present locally. In phase 2, 100 plants in flower were placed adjacent to the focal patch of flowerless plants. Colors indicate individual bee colonies. Damage by bees was significantly higher in phase 1 [generalized linear mixed model fit by maximum likelihood, term = phase, *df* = 1, *F* = 712.70, *P* < 0.001, Akaike information criterion (AIC) 1660.9]. Dots indicate days on which weather prevented data collection. The black line represents a 7-day centered moving average of new bee holes per day. Six flowerless plant species, 36 flowerless plants per bee colony, and five bee colonies (180 flowerless plants in total). (B) 2019 experiment with queenright colonies. Colonies in both treatments were placed adjacent to focal patches of flowerless plants. Colonies in the flowering treatment also had access to 30 plants in flower, placed adjacent to the focal patch, as well as a rooftop garden planted with wildflowers. On 2 days, indicated by triangles, all colonies were given access to sugar solution for 24 hours to mitigate effects of adverse weather conditions. Damage by bees was significantly higher in the flowerless treatment (GLM fit by maximum likelihood, term = treatment, *df* = 1, *F* = 5217.7, *P* < 0.001, AIC 2049.3). Damage levels on roof 2 increased significantly in the month after the wildflower garden was mown (*df* = 30, *P* ≤ 0.001, AIC 2960.4, GLM). Seven flowerless plant species, 600 flowerless plants per treatment, and eight bee colonies per treatment.



queenright colonies and tracked damage by multiple replicate colonies that were simultaneously exposed to environments differing in floral resource availability. The 2019 study made use of the same rooftop employed in 2018 (roof 1) as well as a second nearby rooftop (~200 m from the first). Eight queenright *B. terrestris* hives were placed on each roof, adjacent to focal patches comprising 300 flowerless plants (from seven different species; see supplementary materials for details), which were replaced after 3 weeks. No plants in flower were present on roof 1. On roof 2, bees had access to a preexisting rooftop garden planted with wildflowers (measuring ~4.5 m by 7 m, located ~20 m from the focal patch) as well as a patch of 30 plants in flower (*Antirrhinum* sp.) placed adjacent to the focal patch of flowerless plants. This experiment was initiated in late May, and we monitored damage on 150 randomly chosen plants from each focal patch. At the end of June, the wildflower garden was mown and other plants in flower were removed from roof 2 (see supplementary materials for detailed methods). Bees damaged plants in both focal patches; however, damage levels were consistently higher on roof 1, where no floral resources were present. Furthermore, damage levels on roof 2 increased significantly in the month after the wildflower garden was mown (before tapering off on both roofs as hives approached their reproductive switch points). These results provide further evidence that damage behavior is influenced by the availability of floral resources, complementing the findings from our 2018 rooftop study as well as our pollen-deprivation laboratory experiments.

Taken together, the results of our laboratory and rooftop experiments show that (i) bumble bees damage the leaves of flowerless plants in a way that accelerates flowering time; (ii) the availability of pollen influences this behavior; (iii) the behavior persists under seminatural conditions where bees have the option of foraging farther afield; and (iv) wild bees, in-

cluding species other than *B. terrestris*, damage plant leaves. These findings thus establish that leaf-damaging behavior is a notable component of bumble bee behavior that can influence the local availability of floral resources. The current findings do not establish the mechanism by which bee damage accelerates flowering or the relationship of the observed response to previously documented pathways involved in stress-induced flowering (23). The fact that mechanical damage alone does not fully reproduce the observed effect raises the possibility that some additional bee-derived cue may be at play.

Flower scarcity poses serious challenges for pollinator populations (26), and bumble bees have a particularly pressing need for floral resources during early spring, when queens emerging from diapause must establish their colonies (27). Damaging plant leaves when floral resources are scarce might therefore be an adaptive strategy for accelerating flower production. Conversely, because pollinators are powerful agents of plant evolution (28, 29), it may also be adaptive for some plant species to actively respond to bee-inflicted damage or associated cues (30) to mitigate asynchrony. Furthermore, the demonstration that bee-inflicted leaf damage can have strong effects on time to flowering may have important ecological implications, including for the resilience of plant-pollinator interactions to increases in phenological asymmetry caused by anthropogenic environmental changes.

REFERENCES AND NOTES

1. T. Benton, *Bumblebees: The Natural History and Identification of the Species Found in Britain*, vol. 98 of *Collins New Naturalist Library* (Harper UK, 2006).
2. W. Fagan *et al.*, *Theor. Ecol.* **7**, 289–297 (2014).
3. E. E. Crone, N. M. Williams, *Ecol. Lett.* **19**, 460–468 (2016).
4. D. Goulson, E. Nicholls, C. Botías, E. L. Rotheray, *Science* **347**, 1255957 (2015).
5. M. Schenk, J. Krauss, A. Holzschuh, *J. Anim. Ecol.* **87**, 139–149 (2018).
6. J. T. Kerr *et al.*, *Science* **349**, 177–180 (2015).
7. J. Memmott, P. G. Craze, N. M. Waser, M. V. Price, *Ecol. Lett.* **10**, 710–717 (2007).
8. N. E. Miller-Struttmann *et al.*, *Science* **349**, 1541–1544 (2015).

9. C. Carvell *et al.*, *Nature* **543**, 547–549 (2017).
10. J. D. Thomson, *Philos. Trans. R. Soc. London Ser. B* **365**, 3187–3199 (2010).
11. J. Settele, J. Bishop, S. G. Potts, *Nat. Plants* **2**, 16092 (2016).
12. T. H. Roulston, K. Goodell, *Annu. Rev. Entomol.* **56**, 293–312 (2011).
13. G. Kudo, Y. Nishikawa, T. Kasagi, S. Kosuge, *Ecol. Res.* **19**, 255–259 (2004).
14. F. Requier, K. K. Jowanowitsch, K. Kalinik, I. Steffan-Dewenter, *Ecology* **101**, e02946 (2020).
15. J. Forrest, J. Thomson, *Ecol. Monogr.* **81**, 469–491 (2011).
16. R. J. Stelzer, L. Chittka, M. Carlton, T. C. Ings, *PLOS ONE* **5**, e9559 (2010).
17. P. Willmer, *Curr. Biol.* **22**, R131–R132 (2012).
18. I. Bartomeus *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 20645–20649 (2011).
19. J. E. Ogilvie *et al.*, *Ecol. Lett.* **20**, 1507–1515 (2017).
20. C. Parmesan, *Annu. Rev. Ecol. Syst.* **37**, 637–669 (2006).
21. M. Mikát, C. Franchino, S. Rehan, *Behav. Ecol. Sociobiol.* **71**, 135 (2017).
22. L. D. Harder, *Oecologia* **85**, 41–47 (1990).
23. K. Takeno, *J. Exp. Bot.* **67**, 4925–4934 (2016).
24. N. P. Lemoine, D. Doublet, J.-P. Salminen, D. E. Burkepile, J. D. Parker, *Ecology* **98**, 1817–1828 (2017).
25. P. Smeets, M. Duchateau, *Apidologie* **34**, 333–337 (2003).
26. P. Williams, J. Osborne, *Apidologie* **40**, 367–387 (2009).
27. R. Mallinger, J. Gibbs, C. Gratton, *Landsc. Ecol.* **31**, 1523–1535 (2016).
28. D. D. Gervasi, F. P. Schiestl, *Nat. Commun.* **8**, 14691 (2017).
29. A. A. Agrawal, A. P. Hastings, M. T. J. Johnson, J. L. Maron, J.-P. Salminen, *Science* **338**, 113–116 (2012).
30. M. C. Mescher, C. M. De Moraes, *J. Exp. Bot.* **66**, 425–433 (2015).
31. F. G. Pashalidou, H. Lambert, T. Peybernes, M. Mescher, C. De Moraes, Bumble bees damage plant leaves and accelerate flower production when pollen is scarce, Version 2, Dryad (2020); <https://doi.org/10.5061/dryad.9ghx3fddv>.

ACKNOWLEDGMENTS

The authors thank L. Eyman and G. Ulrich for help with experiments, G. Losapio for statistical advice, and H. Pulido for photography.

Funding: This research was funded by ETH Zürich, Switzerland.

Author contributions: F.G.P., H.L., M.C.M., and C.M.D.M. designed the research. F.G.P., H.L., and T.P. carried out the research. F.G.P., H.L., T.P., M.C.M., and C.M.D.M. interpreted the data. F.G.P., H.L., M.C.M., and C.M.D.M. wrote the paper. All authors gave comments and final approval for publication. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the manuscript, the supplementary materials, or at Dryad (31).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6493/881/suppl/DC1
Materials and Methods
Figs. S1 to S5
Tables S1 to S3
References (32, 33)
Movie S1

17 May 2019; resubmitted 31 January 2020

Accepted 10 April 2020

10.1126/science.aay0496

CHEMICAL PHYSICS

Simultaneous observation of nuclear and electronic dynamics by ultrafast electron diffraction

Jie Yang^{1,2,*}, Xiaolei Zhu^{1,2,3†}, J. Pedro F. Nunes⁴, Jimmy K. Yu^{2,3,5}, Robert M. Parrish^{1,2,3}, Thomas J. A. Wolf^{1,2}, Martin Centurion⁴, Markus Gühr⁶, Renkai Li^{1‡}, Yusong Liu⁷, Bryan Moore⁴, Mario Niebuhr⁶, Suji Park^{1§}, Xiaozhe Shen¹, Stephen Weathersby¹, Thomas Weinacht⁷, Todd J. Martinez^{1,2,3*}, Xijie Wang^{1*}

Simultaneous observation of nuclear and electronic motion is crucial for a complete understanding of molecular dynamics in excited electronic states. It is challenging for a single experiment to independently follow both electronic and nuclear dynamics at the same time. Here we show that ultrafast electron diffraction can be used to simultaneously record both electronic and nuclear dynamics in isolated pyridine molecules, naturally disentangling the two components. Electronic state changes ($S_1 \rightarrow S_0$ internal conversion) were reflected by a strong transient signal in small-angle inelastic scattering, and nuclear structural changes (ring puckering) were monitored by large-angle elastic diffraction. Supported by *ab initio* nonadiabatic molecular dynamics and diffraction simulations, our experiment provides a clear view of the interplay between electronic and nuclear dynamics of the photoexcited pyridine molecule.

Nonadiabatic processes exhibit a complex interplay between the electronic and nuclear degrees of freedom. For electronically excited polyatomic molecules, the vibronic or nonadiabatic couplings become so strong that the Born-Oppenheimer approximation often fails (1). This mixing between electronic and nuclear degrees of freedom presents great challenges to common experimental and theoretical approaches. Specifically, it is experimentally challenging to measure both electronic and nuclear dynamics independently within a single experiment. For time-resolved measurement of photoexcited molecules, most pump-probe spectroscopy focuses on the population dynamics between electronic states. Valence electron spectroscopy could be sensitive to both electron and nuclear dynamics, but the two contributions are often difficult to disentangle (2). Core electron spectroscopy, such as extended x-ray absorption fine structure, can resolve the structural dynamics around a local site but typically requires the presence of one or more heavy atoms (3). Time-resolved diffraction (TRD)

techniques, including ultrafast electron diffraction (UED) pioneered by Zewail and colleagues in the 1990s (4) and time-resolved x-ray diffraction (TRXD) enabled recently by x-ray free-electron lasers, are able to resolve motion of atomic nuclei during photochemical reactions with femtosecond temporal resolution and sub-angstrom spatial resolution (5–7) but have so far been insensitive to electronic dynamics. In most TRD experiments, the independent atom model (IAM), in which the electron redistribution due to bond formation is completely ignored, is invoked to interpret the diffraction patterns. A recent TRXD study by Stankus *et al.* included the valence electron contribution to the elastic scattering signal in a Rydberg excitation, where the diffraction signature of electronic and nuclear dynamics is intertwined (7). As suggested from theory (8, 9), the inelastic scattering signal is expected to reflect electronic dynamics but, to the best of our knowledge, has yet to be used in a TRD experiment. In this work, we used UED to study the photophysics of pyridine excited to the $S_1(n\pi^*)$ state. We show that the inelastic electron scattering modulates the scattering pattern exclusively at small angles and thus is a sensitive observable related directly to the excited state population. By contrast, elastic scattering signals at higher angles encode geometric information. The clean separation of elastic and inelastic scattering signals enables a single UED experiment to simultaneously resolve both electronic [$S_1 \rightarrow S_0$ internal conversion (IC)] and nuclear (ring-puckering) dynamics of the $S_1(n\pi^*)$ state in pyridine.

Pyridine is one of the simplest heterocyclic compounds with rich nonadiabatic dynamics. The investigation of its photophysics is central to the understanding of interplay between $\pi\pi^*$ and $n\pi^*$ states in heterocyclic compounds,

which is critical for the photoprotection mechanism of nucleobases (10). It has been shown that the radiationless transition of the pyridine $S_1(n\pi^*)$ state is extremely sensitive to excess vibrational energy: Higher excess vibrational energy increases IC and decreases intersystem crossing quantum yield (11). This behavior is reminiscent of the “channel-three” decay in benzene and has drawn considerable interest in past decades (12, 13). Despite being extensively studied, the photophysics of such a simple molecule is still under heated debate. Lim (14) proposed a proximity effect mechanism, in which a pseudo-Jahn-Teller effect between the close-lying $n\pi^*$ and $\pi\pi^*$ states promotes an out-of-plane torsional motion, which then greatly enhances the Franck-Condon factor for $S_1 \rightarrow S_0$ IC. The strong dependence of the $(S_1 \rightarrow S_0)/(S_1 \rightarrow T_1)$ branching ratio on the excess vibrational energy is explained by a higher energy barrier for the IC pathway. This model is supported by experimental absorption, fluorescence, and time-resolved photofragment spectroscopy (13, 15, 16). Sobolewski and Domcke (17) and Chachisvilis and Zewail (18) proposed that a crossing of the $S_2(\pi\pi^*)$ and $S_1(n\pi^*)$ states leads to the formation of a prefulvenic structure. Zhong *et al.* (19) proposed an isomerization to Dewar and Hückel structures after passing through a conical intersection (CI). Lobastov *et al.* (20) and Srinivasan *et al.* (21) reported ring opening of pyridine as the dominant pathway after excitation with 267-nm light. These conflicting models, each with its own experimental evidence, persist in part because none of the previous experiments directly measured the electronic and structural dynamics independently and simultaneously.

In most UED experiments, data are analyzed with the IAM, which ignores all electronic redistribution due to chemical bonding or electronic excitation. Specifically, this model neglects two effects: the binding effect that comes from the redistribution of average electron density due to bond formation (one-electron effect) and the correlation effect that comes from electrons avoiding each other because of Coulomb repulsion and Pauli exclusion (two-electron effect) (22–27). Iijima *et al.* (22) and Bartell and Gavin (23) showed that the binding and correlation effects exclusively contribute to elastic and inelastic scattering, respectively (see supplementary materials). Typically, UED experiments use detectors without energy selectivity, and thus elastic and inelastic scattering are recorded together without distinction.

Our experimental setup has been introduced previously (5, 6, 28) and is schematically shown in Fig. 1A. Briefly, the 265-nm pump laser and the 3.7-MeV probe electrons intersected the target gas jet almost colinearly, and the overall instrumental response function (IRF) had a full width at half maximum of ~150 fs (28). To

¹SLAC National Accelerator Laboratory, Menlo Park, CA, USA. ²Stanford PULSE Institute, SLAC National Accelerator Laboratory, Menlo Park, CA, USA. ³Department of Chemistry, Stanford University, Stanford, CA, USA. ⁴Department of Physics and Astronomy, University of Nebraska–Lincoln, Lincoln, NE, USA. ⁵Biophysics Program, Stanford University, Stanford, CA, USA. ⁶Institut für Physik und Astronomie, Universität Potsdam, Potsdam, Germany. ⁷Department of Physics and Astronomy, Stony Brook University, Stony Brook, NY, USA.

*Corresponding author. Email: jieyang@slac.stanford.edu (J.Y.); todd.martinez@stanford.edu (T.J.M.); wangxj@slac.stanford.edu (X.W.)

†These authors contributed equally to this work.

‡Present address: Department of Engineering Physics, Tsinghua University, Beijing, China.

§Present address: Center for Functional Nanomaterials, Brookhaven National Laboratory, Upton, NY, USA.

access inelastic scattering at small angles, we intentionally set the main electron beam off-center from the hole of the detector phosphor screen (Fig. 1B). This setup provided access down to $s = 0.3 \text{ \AA}^{-1}$ in one quadrant, where s is the momentum transfer of the scattered electrons. Small-angle electron scattering is dominated by the inelastic component. For $0.3 < s < 1 \text{ \AA}^{-1}$, previous gas electron diffraction experiments have suggested that the inelastic scattering intensity is typically 5 to 10 times as large as the elastic scattering intensity (29). The pump laser launches a wave packet on the $S_1(n\pi^*)$ surface with $\sim 3000\text{-cm}^{-1}$ excess energy, and the molecule relaxes through a CI along a ring-puckering coordinate (14, 15). Figure 1C shows the S_0 and S_1 potential energy surfaces at the floating occupation molecular orbital-complete active space configuration interaction (FOMO-CASCI) level of theory (supplementary materials).

The experimental and simulated UED data are given in Fig. 2. Figure 2A shows the experimental percentage difference (PD) signal, PD^{exp} , defined as

$$\text{PD}(s; t) = \frac{I(s; t) - I(s; t < 0)}{I(s; t < 0)} \times 100 \quad (1)$$

where $I(s; t)$ is the radially averaged diffraction intensity at any pump-probe delay time t , and $I(s; t < 0)$ is a reference pattern taken before the arrival of the pump laser. In PD^{exp} , the signal at $s > 1.1 \text{ \AA}^{-1}$ appeared around $t = 0$ and persisted throughout our observation window, which reflects structural change of the molecule and will be discussed later. The signal at $s < 1.1 \text{ \AA}^{-1}$, however, contained a sharp rise (IRF-limited) and a slower decay ($1.1 \pm 0.2 \text{ ps}$); see Fig. 2F. To understand this signal, we performed dynamics, quantum chemistry and diffraction simulations. The nonadiabatic dynamics of the lowest three singlet and four triplet states, including spin-orbit interactions, was simulated using the generalized ab initio multiple spawning (GAIMS) (30) method (supplementary materials). Diffraction patterns and PD were first calculated using the IAM (hereafter $\text{PD}_{\text{IAM}}^{\text{sim}}$); see Fig. 2B. $\text{PD}_{\text{IAM}}^{\text{sim}}$ captured most of the large-angle features in PD^{exp} but did not capture the strong increase at small angles ($s < 1 \text{ \AA}^{-1}$; $0 < t < \sim 1.5 \text{ ps}$), indicating that the small-angle signal originated from electron dynamics through either binding (elastic) or correlation (inelastic) effects.

To properly account for scattering signal from both binding and correlation effects, we performed ab initio electron diffraction (AIED) simulations on GAIMS trajectories, similar to the method developed by Breitenstein *et al.* (31) and implemented in TeraChem (32) (supplementary materials). The resulting total, elastic, and inelastic PD ($\text{PD}_{\text{Total}}^{\text{sim}}$, $\text{PD}_{\text{Elastic}}^{\text{sim}}$, and

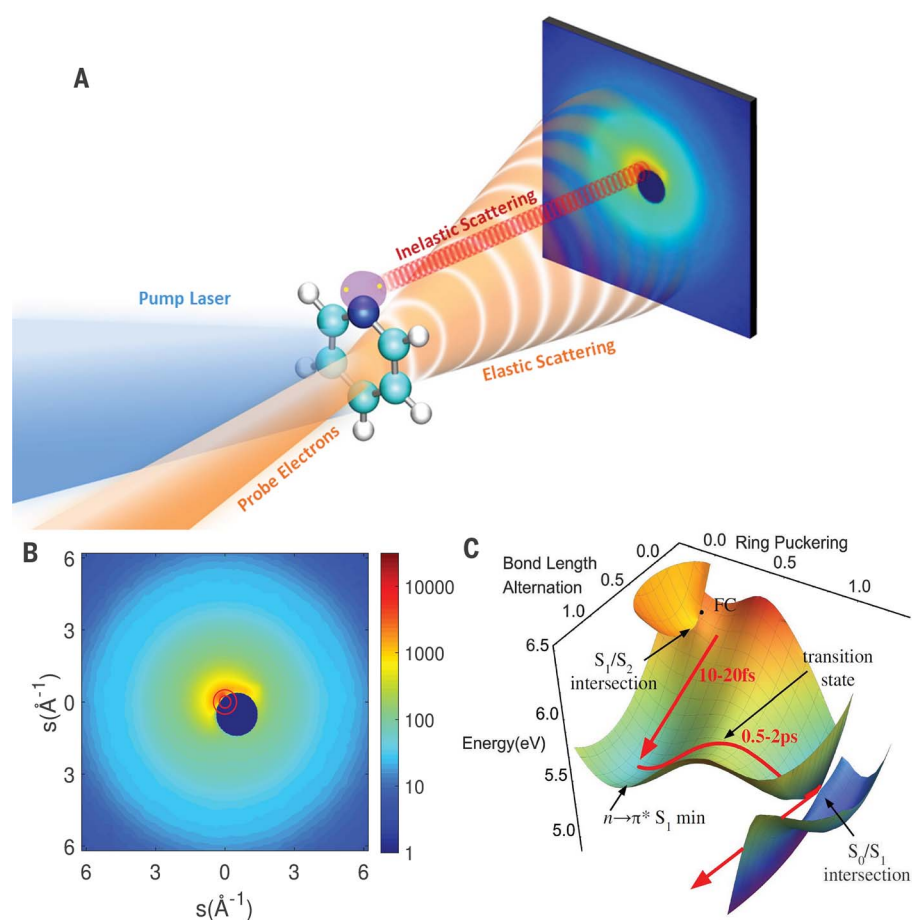


Fig. 1. Experiment overview. (A) Conceptual drawing of the experiment. The inelastic scattering (red scattered wave) concentrates at small angles and encodes information about electron correlation [represented graphically by the two electrons (yellow) in the lone-pair orbital (purple)]. The elastic scattering (orange scattered wave) dominates at high angles and encodes the molecular structure. (B) Diffraction pattern with low- s access. The two red rings near the center represent 0.25 \AA^{-1} (inner ring) and 0.5 \AA^{-1} (outer ring), respectively. (C) Pyridine potential energy surfaces (S_1 and S_0) along the bond length alternation and ring-puckering coordinates (fig. S6). Critical points (FC, Franck-Condon point) and the minimum energy reaction pathway are marked. Time scales were obtained from simulation.

$\text{PD}_{\text{Inelastic}}^{\text{sim}}$) are shown in Fig. 2, C to E. $\text{PD}_{\text{Elastic}}^{\text{sim}}$ and $\text{PD}_{\text{Inelastic}}^{\text{sim}}$ are defined as

$$\text{PD}_{\text{Elastic}}^{\text{sim}} = \frac{I_{\text{Elastic}}(s; t) - I_{\text{Elastic}}(s; t < 0)}{I(s; t < 0)} \times 100 \quad (2)$$

$$\text{PD}_{\text{Inelastic}}^{\text{sim}} = \frac{I_{\text{Inelastic}}(s; t) - I_{\text{Inelastic}}(s; t < 0)}{I(s; t < 0)} \times 100 \quad (3)$$

The $\text{PD}_{\text{Total}}^{\text{sim}}$ nicely captured all the major features in PD^{exp} , with $\text{PD}_{\text{Elastic}}^{\text{sim}}$ and $\text{PD}_{\text{Inelastic}}^{\text{sim}}$

separately located in the $s > \sim 1.1 \text{ \AA}^{-1}$ and $s < \sim 1.1 \text{ \AA}^{-1}$ regions. Lineout plots of the three strongest features—the peak at $0.3 < s < 0.7 \text{ \AA}^{-1}$, the peak at $3.9 < s < 4.9 \text{ \AA}^{-1}$, and the trough at

$5.5 < s < 6.5 \text{ \AA}^{-1}$ —are shown in Fig. 2, F to H. Figure 2F shows that the IAM simulation fails to reproduce the experimental data at small s and that the AIED simulation and experiment agree reasonably well—both show fast-rising (IRF-limited) and slow-decaying (experiment: $1.1 \pm 0.2 \text{ ps}$, simulation: $1.3 \pm 0.1 \text{ ps}$) features. In addition, Fig. 2, D and E, shows that this signal exclusively came from the inelastic component.

We then inspected the connection between the simulated small-angles scattering signal and the electronic state population. The GAIMS calculation predicted that the $S_1(n\pi^*)$ state was exclusively populated and that $>90\%$ of the population returned to the ground state S_0 within the 2.4-ps simulation window. The small-angle signal $\text{PD}_{\text{Total}}^{\text{sim}}(0.3 < s < 0.7 \text{ \AA}^{-1}; t)$ correlates well with the S_1 population in the GAIMS simulation, as illustrated in Fig. 3A.

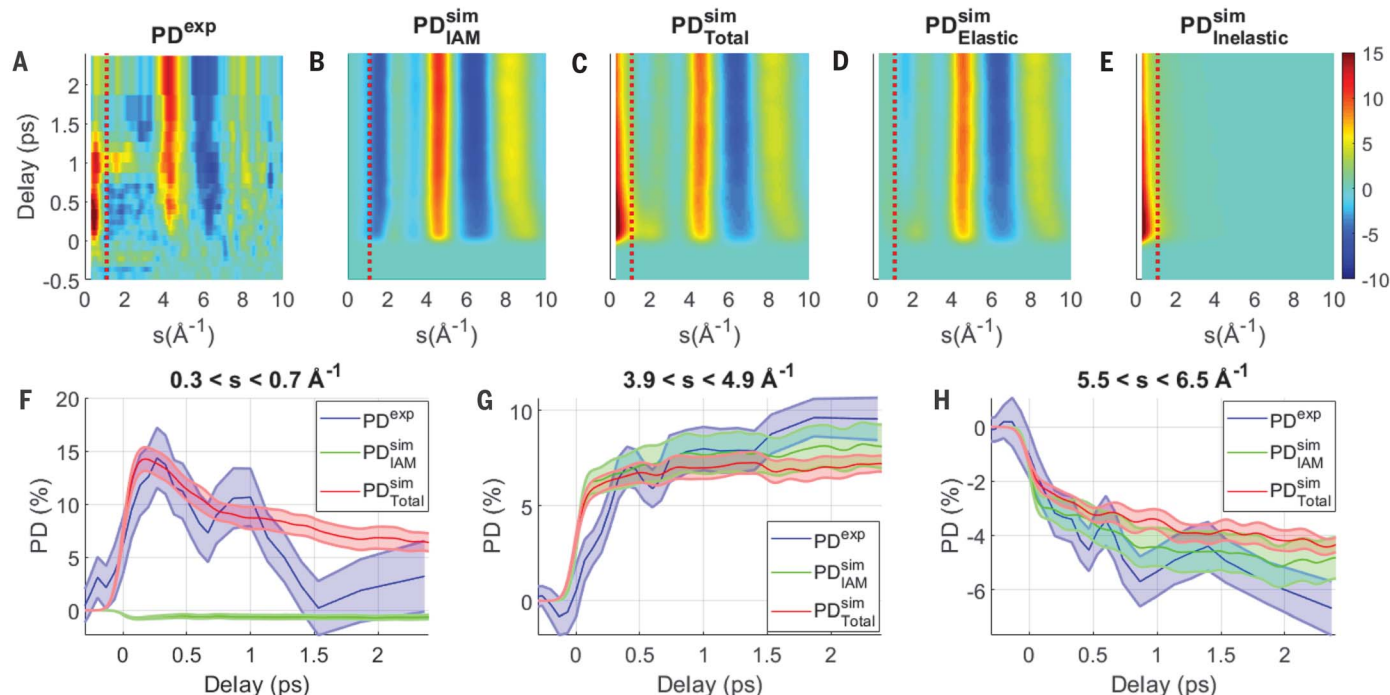
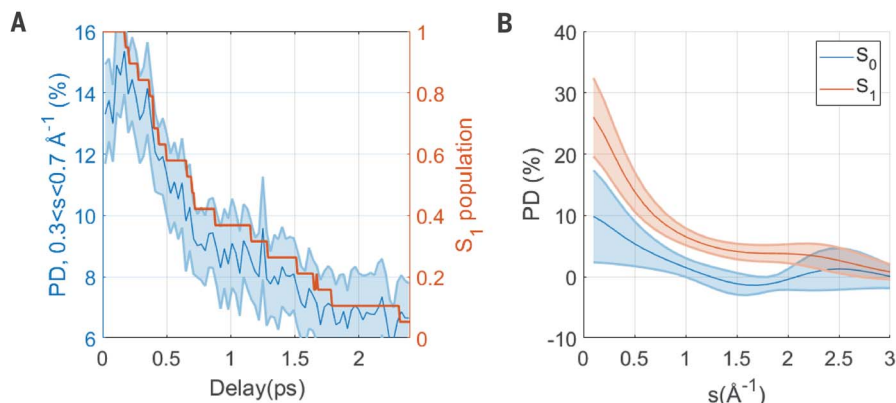


Fig. 2. Experimental and simulated UED signal for pyridine. (A) Experimental PD signal, normalized by the 9% excitation ratio. (B) Simulated PD signal from the IAM (elastic component only), using all 24 initial and 2040 spawned trajectories. (C) Simulated PD total (elastic and inelastic) signal using AIED. (D) Elastic part of simulated PD using AIED. (E) Inelastic part of simulated PD using AIED. In (A) to (E), the red dotted line denotes $s = 1.1 \text{ \AA}^{-1}$.

(F to H) Lineout plots of the three strongest features from the experiment, plotted together with IAM and AIED simulations. Uncertainties in (F) to (H) are represented by shaded regions, calculated with one SD of a bootstrapped dataset (experiment) or one SEM of all trajectories (simulation). Simulation results are convolved with a 150-fs Gaussian kernel to account for the experimental IRF.

Fig. 3. Small s scattering and excited state population. (A) Simulated PD signal using AIED (blue, left y axis) and S_1 population (red, right y axis). (B) The $PD^{\text{sim}}_{\text{Total}}$ signal on S_0 and S_1 integrated over the time course of the simulation (0 to 2.4 ps), calculated using all accessed geometries by each state. Uncertainty is represented by shaded regions, calculated with one SEM of all trajectories (A) or one SD of all visited geometries (B).



To further show this correlation, we simulated the $PD^{\text{sim}}_{\text{Total}}$ signal for S_0 and S_1 states over all trajectories (Fig. 3B). We found that the S_1 signal was about three times as strong as the S_0 signal at $0.3 < s < 1 \text{ \AA}^{-1}$. This simulation therefore confirms that the small-angle signal could be used to trace the excited state population, and the experimentally observed positive signal ($s < 1 \text{ \AA}^{-1}$; $0 < t < \sim 1.5 \text{ ps}$) shows both the $S_0 \rightarrow S_1$ photoexcitation and the $S_1 \rightarrow S_0$ IC.

Here we provide a simple physical picture for the observed inelastic signal. The diffraction signature of the correlation effect is a

negative contribution to small-angle inelastic scattering, and most (80 to 90%) of this effect comes from dynamic correlation due to Coulomb repulsion (26, 27, 31). In the ground state, the two electrons in the lone-pair orbital are spatially close and the dynamic correlation is strong. In the photoexcited ($\pi\pi^*$) state, however, the two electrons no longer occupy the same molecular orbital, leading to a marked reduction of the dynamic correlation. According to the above-mentioned arguments, small-angle inelastic scattering is therefore substantially increased by photoexcitation and

decreased upon relaxation to the electronic ground state. We have also simulated the expected AIED signal for the S_2 , T_1 , and T_2 states (fig. S2), which confirms that all of these open-shell excited states give rise to a strong increase in small-angle inelastic scattering. Although we believe this picture provides an intuitive explanation of the observed inelastic signal, the concrete details and generality are subject to future studies.

To extract the nuclear structural dynamics, we used $1.1 < s < 10.5 \text{ \AA}^{-1}$ data to apply a genetic χ^2 structural fitting algorithm (supplementary

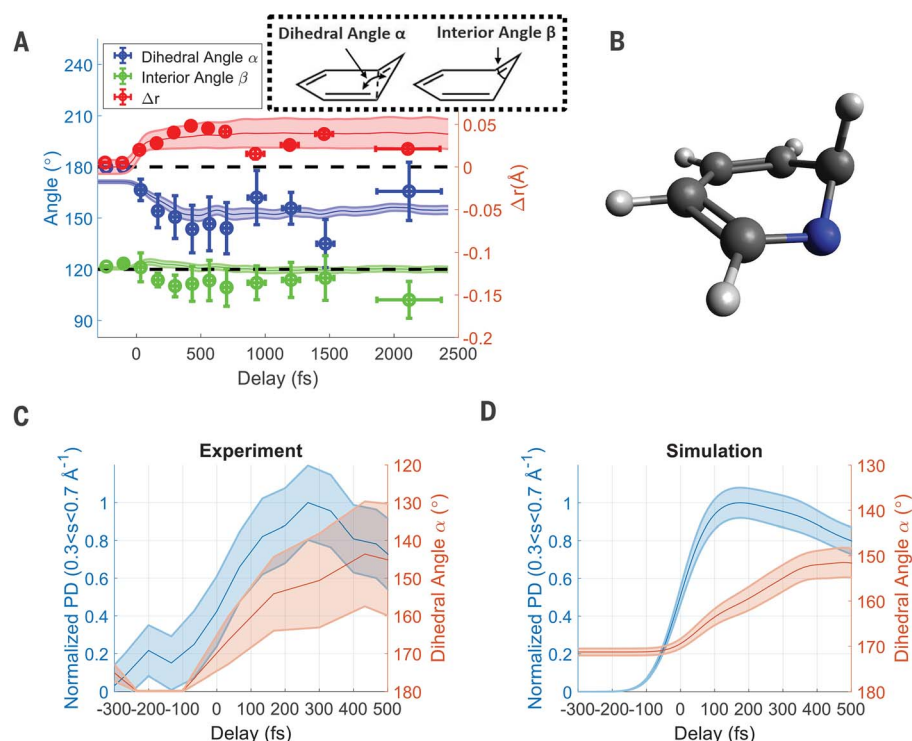


Fig. 4. Combined structural and electronic dynamics. (A) Experimental structural evolution (circles with error bars) retrieved from a genetic χ^2 fitting algorithm. Vertical error bars represent one SD of 86 individual fitting results; horizontal error bars represent the time window used for each fitting. Simulated values (lines with shaded regions) were obtained from weighted averages of all spawned trajectories from GAIMS simulation. The two dashed lines show the position of 120° and 180° . (B) Geometry of the most-accessed S_1/S_0 CI from GAIMS simulation. Blue, nitrogen atom. (C and D). Experimental (C) and simulated (D) temporal evolution of small-angle PD signal (blue, normalized to the maximum value) and dihedral angle (red). Uncertainty is represented by shaded regions, calculated by one SD of a bootstrapped dataset (experiment) or one SEM of all trajectories (simulation). Simulation results are convolved with a 150-fs Gaussian kernel to account for experimental IRF.

materials) (33). Three fitted structural parameters—dihedral angle α , interior angle β , and average bond length of the ring Δr —are shown in Fig. 4A (circles with error bars), together with the value extracted from the GAIMS simulation (lines with shaded regions). Because the contrast between a nitrogen atom and a CH group is small in electron diffraction, it is difficult to experimentally distinguish which atom puckers out of the plane. Therefore, the dihedral angle α and interior angle β are both referring to the heavy atom with the largest out-of-plane torsion (Fig. 4A, inset). Both experiment and simulation show a large ($\sim 30^\circ$) change in dihedral angle and a ~ 0.04 -Å expansion of the ring bond lengths, with a small change in angle β . The most-accessed S_1/S_0 CI geometry from the GAIMS simulation is shown in Fig. 4B. It is described by $\sim 60^\circ$ out-of-plane torsion on a carbon atom adjacent to the nitrogen atom. Because the molecule spent relatively little time at the CI and the torsion angle was smaller both before and after the CI (Fig. 1B), the averaged out-of-plane torsion at any time appeared to be only $\sim 30^\circ$.

Taking into account all of the analysis discussed above, we were finally able to plot electronic and nuclear dynamics together in Fig. 4, C and D. The small-angle PD signal at $0.3 < s < 0.7 \text{ Å}^{-1}$ represents excited state population (blue curves), whereas the dihedral angle represents the structural change along the main reaction coordinate (red curves). The torsion started ~ 100 fs after the $S_0 \rightarrow S_1$ photoexcitation, agreeing well with the model that

a small barrier is present in the out-of-plane ring-puckering coordinate. The $S_1 \rightarrow S_0$ IC started only after the dihedral angle reached a certain point ($\sim 150^\circ$ for experiment and $\sim 160^\circ$ for simulation), consistent with the prediction that the $S_1 \rightarrow S_0$ IC requires a relatively large out-of-plane torsion. This plot, along with the lack of structural changes in other degrees of freedom, confirms that the out-of-plane torsion was the major motion that drove the $S_1 \rightarrow S_0$ IC. Even though our data are consistent with the UED data reported by Srinivasan *et al.* (21), we have a different interpretation of the dynamics that does not include ring opening. A detailed comparison between the two experiments is given in the supplementary materials.

In summary, through the correlation of a specific nuclear degree of freedom to electronic structure change, we have demonstrated that structural and electronic dynamics can be retrieved simultaneously and independently from a single UED dataset. This method allows us to identify the relaxation mechanism in the $n\pi^*$ state of pyridine. Owing to the universality of the diffraction signature from correlation effects (24–27, 31), we expect that this method will be widely applicable in ultrafast photochemistry. Moreover, the inelastic electron scattering is a Fourier transform of the change of the two-electron density caused by electron correlation. In contrast to spectroscopic probes, it measures spatial rather than energetic aspects of electron correlation (23). Because electron correlation is at the heart of modern quantum chemistry simulations, in-

elastic electron scattering provides a benchmark for state-of-the-art and future theoretical and computational methods.

REFERENCES AND NOTES

- W. Domcke, D. Yarkony, H. Köppel, *Conical Intersections: Electronic Structure, Dynamics & Spectroscopy* (World Scientific, 2004).
- A. Stolow, A. E. Bragg, D. M. Neumark, *Chem. Rev.* **104**, 1719–1757 (2004).
- M. A. Newton, A. J. Dent, J. Evans, *Chem. Soc. Rev.* **31**, 83–95 (2002).
- J. C. Williamson, J. M. Cao, H. Ihee, H. Frey, A. H. Zewail, *Nature* **386**, 159–162 (1997).
- J. Yang *et al.*, *Science* **361**, 64–67 (2018).
- T. J. A. Wolf *et al.*, *Nat. Chem.* **11**, 504–509 (2019).
- B. Stankus *et al.*, *Nat. Chem.* **11**, 716–721 (2019).
- M. Kowalewski, K. Bennett, S. Mukamel, *Struct. Dyn.* **4**, 054101 (2017).
- M. Simmermacher, A. Moreno Carrascosa, N. E. Henriksen, K. B. Møller, A. Kirrander, *J. Chem. Phys.* **151**, 174302 (2019).
- R. Improta, F. Santoro, L. Blancafort, *Chem. Rev.* **116**, 3540–3593 (2016).
- I. Yamazaki, K. Sushida, H. Baba, *J. Chem. Phys.* **71**, 381–387 (1979).
- C. E. Otis, J. L. Knee, P. M. Johnson, *J. Phys. Chem.* **87**, 2232–2239 (1983).
- B. X. Wang, B. K. Liu, Y. Q. Wang, L. Wang, *Int. J. Mass Spectrom.* **289**, 92–97 (2010).
- E. C. Lim, *J. Phys. Chem.* **90**, 6770–6777 (1986).
- E. Villa, A. Amirav, E. C. Lim, *J. Phys. Chem.* **92**, 5393–5397 (1988).
- J. I. Selco, P. L. Holt, R. B. Weisman, *J. Chem. Phys.* **79**, 3269–3278 (1983).
- A. L. Sobolewski, W. Domcke, *Chem. Phys. Lett.* **180**, 381–386 (1991).
- M. Chachisvilis, A. H. Zewail, *J. Phys. Chem. A* **103**, 7408–7418 (1999).
- D. P. Zhong *et al.*, *Chem. Phys. Lett.* **298**, 129–140 (1998).
- V. A. Lobastov *et al.*, *J. Phys. Chem. A* **105**, 11159–11164 (2001).
- R. Srinivasan, J. S. Feenstra, S. T. Park, S. Xu, A. H. Zewail, *Science* **307**, 558–563 (2005).
- T. Iijima, R. A. Bonham, T. Ando, *J. Phys. Chem.* **67**, 1472–1474 (1963).
- L. S. Bartell, R. M. Gavin, *J. Am. Chem. Soc.* **86**, 3493–3498 (1964).

24. J. H. Wang, A. N. Tripathi, V. H. Smith Jr., *J. Chem. Phys.* **101**, 4842–4854 (1994).
25. Y. Sasaki, H. Takeuchi, S. Konaka, M. Kimura, *Int. J. Quantum Chem.* **43**, 701–712 (1992).
26. J. J. McClelland, M. Fink, *Phys. Rev. Lett.* **54**, 2218–2221 (1985).
27. M. Breitenstein, A. Endesfelder, H. Meyer, A. Schweig, W. Zittlau, *Chem. Phys. Lett.* **97**, 403–409 (1983).
28. X. Shen et al., *Struct. Dyn.* **6**, 054305 (2019).
29. S. Shibata, F. Hirota, N. Kakuta, T. Muramatsu, *Int. J. Quantum Chem.* **18**, 281–285 (1980).
30. B. F. E. Curchod, C. Rauer, P. Marquetand, L. González, T. J. Martínez, *J. Chem. Phys.* **144**, 101102 (2016).
31. M. Breitenstein, R. J. Mawhorter, H. Meyer, A. Schweig, *Phys. Rev. Lett.* **53**, 2398–2401 (1984).
32. I. S. Ufimtsev, T. J. Martínez, *J. Chem. Theory Comput.* **5**, 2619–2628 (2009).
33. S. Habershon, A. H. Zewail, *ChemPhysChem* **7**, 353–362 (2006).
34. J. Yang, Figures for “Simultaneous Observation of Nuclear and Electronic Dynamics by Ultrafast Electron Diffraction,” Version 1, Zenodo (2020); <http://doi.org/10.5281/zenodo.3732838>.

ACKNOWLEDGMENTS

We thank G. M. Stewart from SLAC National Accelerator Laboratory for assistance in making Fig. 1A. M.G. serves on the Science Advisory Committee of LCLS, which manages the SLAC UED facility. **Funding:** The experimental part of this research was performed at SLAC MeV-UED, which is supported in part by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences (DOE BES) SUF Division Accelerator & Detector R&D program, the LCLS facility, and SLAC under contract nos. DE-AC02-05-CH11231 and DE-AC02-76SF00515. T.J.M., T.J.A.W., X.Z., J.K.Y., and R.M.P. were supported by the AMOS program within the Chemical Sciences, Geosciences, and Biosciences Division of the DOE BES. M.G. and M.N. are funded via a Lichtenberg Professorship of the Volkswagen Foundation. J.P.F.N., B.M., and M.C. were supported by the DOE BES under award no. DE-SC0014170. Y.L. and T.W. were supported by the DOE under award no. DE-FG02-08ER15984. **Author contributions:** J.Y., J.P.F.N., T.J.A.W., Y.L., R.L., S.P., X.S., S.W., T.W., and X.W. carried out the experiments. J.P.F.N., M.N., M.C., and M.G. improved the experimental system. J.P.F.N., B.M., and M.C. tested the sample delivery system. J.Y. analyzed the experimental data and performed the genetic structural retrieval algorithm. X.Z. and T.J.M. performed the GAIMS simulation.

X.Z., R.M.P., and T.J.M. developed the computational method for AIED simulation. X.Z., J.K.Y., and T.J.M. carried out the AIED simulation. J.Y., X.Z., M.G., T.J.M., and X.W. prepared the manuscript, with discussion and improvements from all authors. **Competing interests:** None declared. **Data and materials availability:** All data underlying the figures are deposited at Zenodo (34). The raw experimental data are archived at the SLAC MeV-UED facility, and the raw simulated data are stored at the Martinez laboratory at Stanford University and SLAC. All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6493/885/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S14
Tables S1 and S2
References (35–53)
Movie S1

8 February 2020; accepted 3 April 2020
10.1126/science.abb2235

METROLOGY

Coherent optical clock down-conversion for microwave frequencies with 10^{-18} instability

Takuma Nakamura^{1,2*}, Josue Davila-Rodriguez¹, Holly Leopardi^{1,2,†}, Jeff A. Sherman¹, Tara M. Fortier^{1,2}, Xiaojun Xie³, Joe C. Campbell³, William F. McGrew^{1,2}, Xiaogang Zhang^{1,2}, Youssef S. Hassan^{1,2}, Daniele Nicolodi^{1,2}, Kyle Beloy¹, Andrew D. Ludlow^{1,2}, Scott A. Diddams^{1,2}, Franklyn Quinlan^{1,2*}

Optical atomic clocks are poised to redefine the Système International (SI) second, thanks to stability and accuracy more than 100 times better than the current microwave atomic clock standard. However, the best optical clocks have not seen their performance transferred to the electronic domain, where radar, navigation, communications, and fundamental research rely on less stable microwave sources. By comparing two independent optical-to-electronic signal generators, we demonstrate a 10-gigahertz microwave signal with phase that exactly tracks that of the optical clock phase from which it is derived, yielding an absolute fractional frequency instability of 1×10^{-18} in the electronic domain. Such faithful reproduction of the optical clock phase expands the opportunities for optical clocks both technologically and scientifically for time dissemination, navigation, and long-baseline interferometric imaging.

Motivated by placing tighter constraints on physical constants and their possible variations (1, 2), precise measurements of gravitational potential (3), and gravitational wave detection (4), the accuracy and stability of optical clocks have continued to improve, such that they now outperform their microwave counterparts by orders of magnitude. These clocks—based on optical transitions in atoms and ions such as ytterbium, strontium, and aluminum (5, 6)—take advantage of an operating frequency that can exceed 1000 THz. This corresponds to subdividing a second into mere femtoseconds, allowing for an extremely precise measurement of time. The frequency stability of an optical clock, best described as the fraction of the clock's frequency fluctuations relative to its nominal operating frequency, can now reach below 10^{-18} (3). Perhaps more importantly, optical clocks can reach 10^{-16} performance in a matter of seconds, rather than the month-long averaging required of a microwave Cs fountain clock, which currently defines the Système International (SI) second, to reach this level (7). With such extraordinary performance, in conjunction with optical clocks meeting other benchmarks laid out by the International Committee for Weights and Measures (CIPM) (8), a redefinition of the SI second appears inevitable.

The range of applications enjoyed by optical clocks can be further extended by transferring

their stability to the electronic domain. Doppler radar sensitivity, particularly for slow-moving objects, is strongly determined by the frequency noise of the transmitted microwaves and could see a large sensitivity enhancement by using optically derived electrical signals (9–11). Astronomical imaging and precise geodesy with very-long-baseline interferometry (VLBI) also rely on highly frequency-stable electronic sources (12). In ground-based VLBI, microwave and millimeter wave signals are detected at receivers spread across the globe and are coherently combined to form exceptionally high-resolution images of cosmic objects. Moving to a space-based VLBI network greatly increases the resolution (13) and avoids atmospheric distortions that limit the observation time. In spaced-based VLBI, maintaining phase coherence with electronic local oscillators that have optical clock-level stability could increase the observation time from seconds to hours, with a commensurate increase in the number of objects that can be imaged with high fidelity.

With the use of fiber-based optical frequency combs (OFCs) and state-of-the-art photodetectors, we have generated and evaluated microwave signals that preserve the phase of the optical clocks from which they are derived with subfemtosecond precision. The resulting frequency stability on a 10-GHz carrier is better than any other microwave source and represents a 100-fold stability improvement over the best Cs fountain clocks. Moreover, the inaccuracy of the optical-to-microwave frequency conversion was measured to be less than 1×10^{-19} . Preservation of the optical clock phase opens up the possibility of distant optical clock synchronization with microwave carriers for applications in navigation and fundamental physics. Lastly, coherently linking an optical atomic frequency standard to the electronic domain allows for future calibration of electronic clocks, an important consid-

eration for the redefinition of the SI second based on an optical atomic transition.

Generating an electronic signal linked to an optical clock is the physical implementation of dividing the optical clock frequency by a large integer (10). The concept is shown in Fig. 1. The first element in this division process is the OFC—a laser source consisting of an array of discrete, evenly spaced frequency tones that span hundreds of THz (14, 15). When an OFC is locked to a clock, each individual tone of the comb carries the same frequency stability as that of the master clock. [Transferring the clock stability to each line is nearly perfect; added instabilities are only at the 10^{-20} level or below (16–18).] The broad spectrum of the comb gives rise to a train of optical pulses with subpicosecond duration. The repetition rate of these pulses, typically in the range of tens of MHz to a few GHz, is coherently linked to the optical clock frequency but is divided down to a much lower microwave frequency. Importantly, the clock frequency fluctuations also divide, such that the fractional frequency stability is maintained. Thus, locking an OFC to an optical clock operating at 259 THz and fractional frequency instability of 10^{-16} can produce a 100-MHz pulse train whose repetition fractional frequency instability is also 10^{-16} .

The repetition rate of an OFC is accessible with electronics, such that illuminating a high-speed photodiode with a locked OFC can, in principle, create a train of electrical pulses with optical clock stability. Optical-to-electrical conversion that preserves optical clock-level stability is not straightforward, however, for this process must contend with the photodiode's nonlinear response engendered by the high peak intensities of ultrashort pulses, the quantum limits of light detection, and the vagaries of electron transport dynamics (19–21). Considerable effort has been devoted to understand and overcome the limitations of photodiodes for optical-to-electrical conversion of ultrastable optical pulse trains, leading to new detector designs (22) and techniques to lower the impact of quantum noise in the phase stability of the optically derived electronic signal (23). This progress has set the stage for the demonstration of electrical signal generation that faithfully reproduces the frequency and phase of a state-of-the-art optical clock.

With frequency stability better than any other microwave source, measurements required constructing two systems and comparing them against one another (24). A simplified schematic diagram of the microwave generation and measurement is shown in Fig. 1B. Ten-GHz microwaves were derived from two independent Yb optical lattice clocks, each of which demonstrate state-of-the-art stability (3), and absolute frequency verified against the SI second (25). The OFCs were based on two home-built erbium fiber mode-locked lasers with

¹Time and Frequency Division, National Institute of Standards and Technology, 325 Broadway, Boulder, CO 80305, USA. ²Department of Physics, University of Colorado Boulder, 440 UCB, Boulder, CO 80309, USA. ³Department of Electrical and Computer Engineering, University of Virginia, Charlottesville, VA 22904, USA. ⁴Department of Electrical and Computer Engineering, University of Virginia, Charlottesville, VA 22904, USA.

*Corresponding author. Email: takuma.nakamura@nist.gov (T.N.); franklyn.quinlan@nist.gov (F.Q.)

[†]Present address: Space Dynamics Laboratory, 1695 North Research Park Way, North Logan, UT 84341, USA.

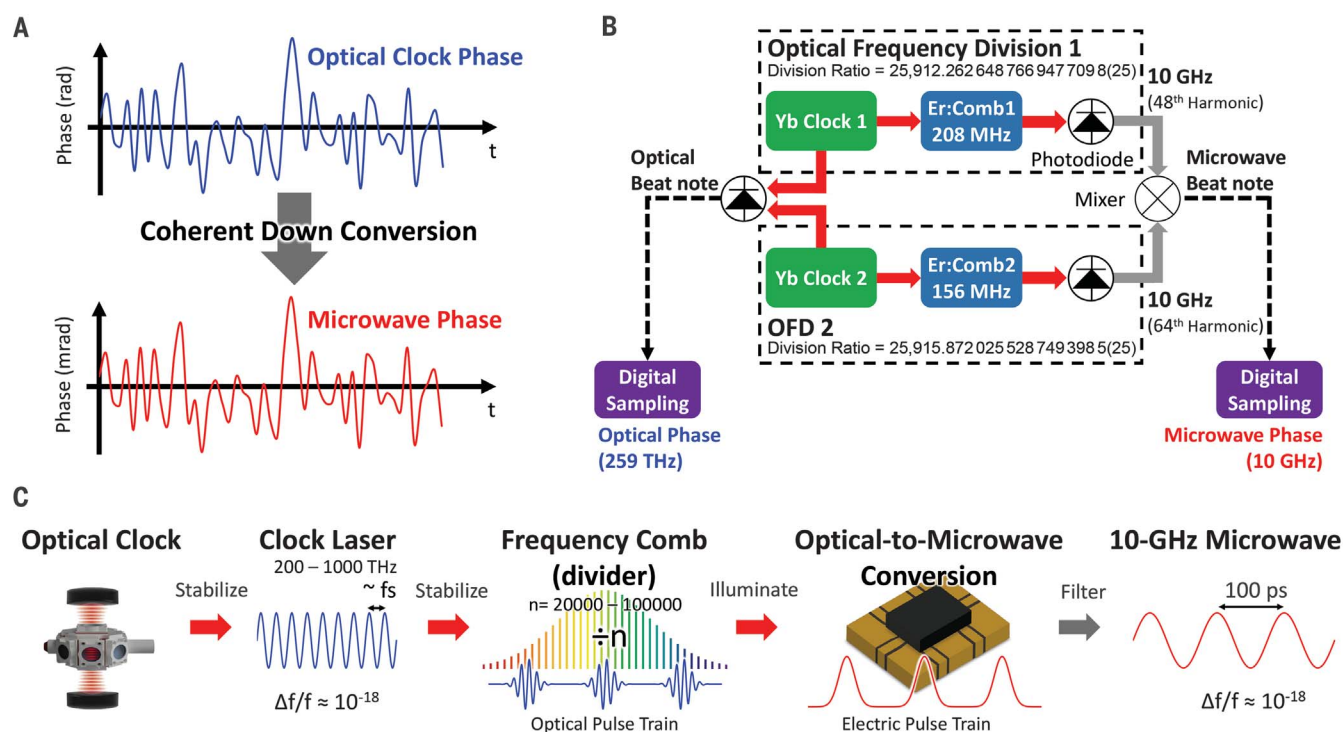


Fig. 1. Coherent optical clock down-conversion. (A) The optical clock phase is transferred to the microwave domain with fluctuations scaled by the optical-to-microwave frequency ratio. (B) Simplified setup of phase and frequency stability measurements. The output of two independent Yb optical atomic clocks generate microwave signals at 10 GHz, where the ratio of the optical to microwave frequencies is determined to 19 digits of precision. By frequency-mixing the 10-GHz outputs, the relative phase fluctuations are

recorded. A direct optical beat note reporting the relative optical phase of the Yb clocks is also recorded. Er, erbium; OFD, optical frequency division. (C) Schematic of microwave generation from an optical atomic clock. An OFC is stabilized to the optical clock laser. Optical-to-microwave conversion through high-speed photodetection generates a train of electrical pulses. Selectively filtering the electrical signal results in a microwave tone that is phase-coherent with the optical clock. $\Delta f/f$, fractional frequency instability.

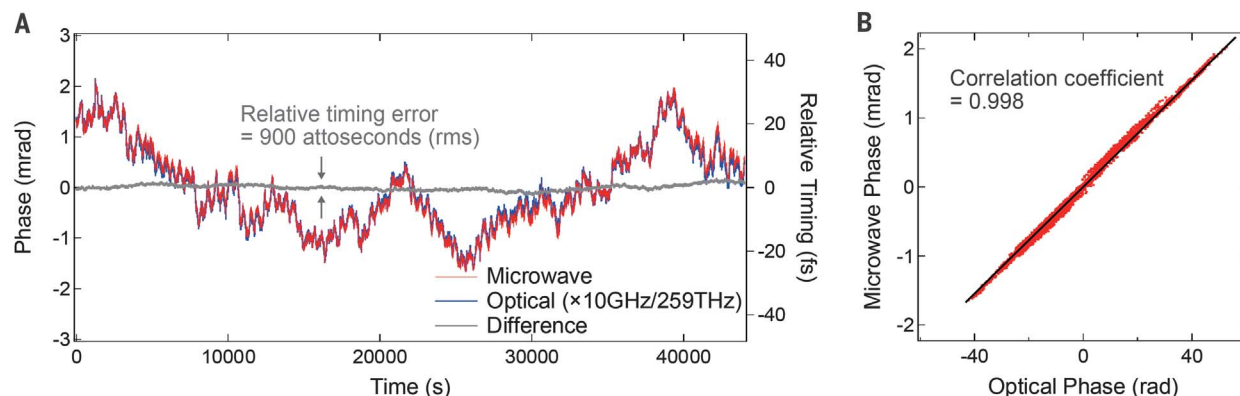


Fig. 2. Optical and microwave phase, timing, and relative coherence. (A) Phase and corresponding timing fluctuations in optical and microwave domains. The point-by-point difference in the (scaled) optical and microwave phase records is shown in gray. (B) Phase correlation plot demonstrating a correlation coefficient of 0.998. The black line is the expected slope given by the optical-to-microwave frequency ratio.

respective repetition rates of 208 and 156 MHz. These OFCs were engineered for long-term, phase-slip-free operation and contribute negligible excess noise. The optical pulse trains from the OFCs were detected with photodiodes designed for high speed and high linearity (22), from which electrical pulse trains were generated. The frequency spectrum of these electri-

cal pulses is an array of tones at the harmonics of the pulse repetition rate. Electrical band-pass filters (100-MHz bandwidth) selected a single frequency near 10 GHz from each system for evaluation. Because the repetition rates of the two lasers are not the same, the nominal 10-GHz outputs represented the 48th harmonic and 64th harmonic of the respective systems.

These 10-GHz outputs were combined in a microwave frequency mixer, producing a difference frequency near 1.5 MHz that was digitally sampled and analyzed with software-defined radio (26), from which the microwave phase was extracted. From the phase, frequency stability and accuracy were determined. In addition to the comparisons of the microwave

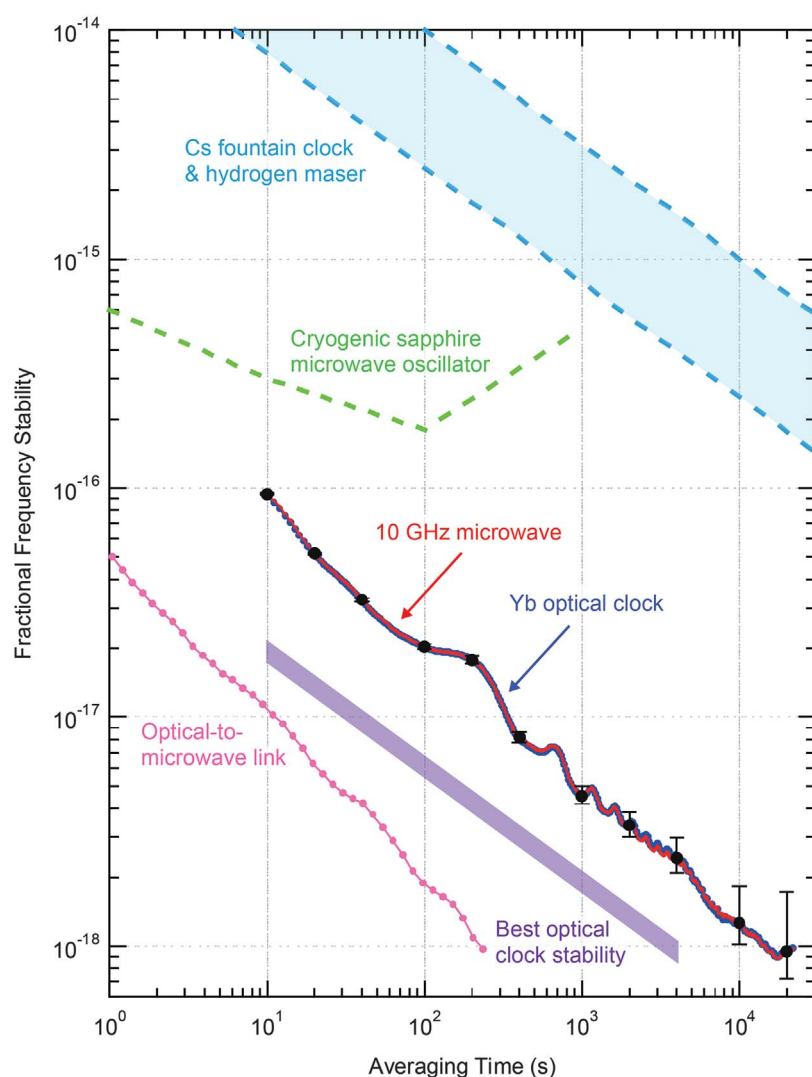


Fig. 3. Fractional frequency stability comparison of state-of-the-art sources. The microwave and optical signals locked to the Yb clocks are reported in terms of the total Allan deviation, with error bars representing 1σ confidence intervals. The white frequency noise asymptote is $1.6 \times 10^{-16} / \sqrt{\tau}$. Additionally, the residual instability of the optical-to-microwave link is well below state-of-the-art optical clocks (7, 29–31). For all plots, the frequency stability is given by the Allan deviation.

outputs, a direct optical comparison of the clocks was made. This was performed by combining the optical clock signals onto a single photodetector, directly generating an electrical signal at an intentionally offset beat frequency between the clocks. This beat frequency was also digitally sampled and recorded. Comparing the phase of the difference frequency of the microwave outputs to that of the optical beat frequency allowed for confirmation of high-fidelity phase and frequency transfer to the electronic domain.

It is interesting to note the level of resolution required to measure a fractional frequency instability of 10^{-16} at 1 s on a 10-GHz signal. This implies tracking phase changes corresponding to only one-millionth of a cycle. As such, standard frequency counting techniques,

although adequate to measure beat frequencies between state-of-the-art optical clocks, cannot yield the required precision for our microwaves. Achieving this level of phase resolution, and maintaining it over several hours, was accomplished in the following ways. First, we utilized microwave amplifiers with low flicker noise, and we routed signals with temperature-insensitive cabling. This gave us an output with ample power (~ 10 mW) without sacrificing stability. Second, by frequency-mixing the 10-GHz outputs, we shifted the measurement to a 1.5-MHz carrier. This reduced the requirements on the fractional stability that we had to measure by ~ 7000 (10 GHz/1.5 MHz). By digitally sampling the 1.5-MHz carrier, the phase difference between the 10-GHz outputs could be tracked with high resolution. Another ad-

vantage of this measurement scheme is that the high phase resolution is achievable without requiring the two sources to oscillate at exactly the same frequency. This gives our measurement system some dexterity in comparing high stability signals from independent sources.

Optical and microwave phase fluctuations, continuously recorded over 44,000 s, are shown in Fig. 2. In Fig. 2A, the phase fluctuations of the optical clock have been scaled by a factor equal to the optical-to-microwave frequency ratio (nearly 26,000) to illustrate the extremely high fidelity in the optical-to-microwave transfer. The relative phase can also be expressed as a timing fluctuation and is bounded by ± 30 fs for both optical and microwave signals. Also shown in Fig. 2A is the point-by-point difference between optical and microwave measurements, limited to root mean square (RMS) fluctuations of 60 mrad, corresponding to a RMS relative timing fluctuation of only 900 as. This implies that optical clocks with even higher stability can be converted to microwave signals without loss of fidelity. The strong correlation between the optical and microwave phases is shown in Fig. 2B. The degree of correlation is quantified by the correlation coefficient, ranging from zero for completely uncorrelated phases to a maximum value of 1 for complete linear correlation between optical and microwave phase (27). The calculated correlation coefficient for the data in Fig. 2 is 0.998. Such femtosecond-level, high coherence optical-to-microwave conversion opens up the possibility of connecting distant optical clocks with a microwave link. Currently, these clocks can be linked optically through fiber or over free space (28), enabling state-of-the-art clock comparisons and synchronization. Free-space optical links are particularly useful for the many situations where a dedicated fiber link is not available but can become ineffective because of poor weather or dusty conditions. The lower loss of microwave transmission could prove advantageous under such conditions by providing a link that would be impossible to maintain optically.

Whereas fluctuations in the phase provide all the frequency and timing stability information of an oscillator, the fractional frequency instability is the more typical performance benchmark. Figure 3 displays the fractional frequency instabilities derived from the same 44,000-s duration phase measurements shown in Fig. 2A. The frequency stability of the derived microwaves followed that of the optical clocks precisely, ultimately yielding an absolute fractional frequency instability of 1×10^{-18} . This is 100 times more stable than the Cs fountain clocks that currently serve as the best realization of the SI second. The short-term stability also exceeds that of other microwave sources, the best of which are microwave oscillators based on whispering gallery mode resonances in cryogenically cooled sapphire (29). There

Table 1. Frequency accuracy evaluation of the optical-to-microwave link. The optical clock frequency offsets were measured both optically and on the derived 10-GHz microwaves, then compared. The difference in the two measurements is consistent with zero at the 10^{-19} level. Yb1, Yb clock 1; Yb2, Yb clock 2.

Measurement	Frequency offset (Yb1 – Yb2) at 259 THz (Hz)	Fractional frequency offset
Optical	–0.0152862	$(-5.8986 \pm 0.095) \times 10^{-17}$
Microwave	–0.0152926	$(-5.9011 \pm 0.096) \times 10^{-17}$
Difference	0.0000064	$(2.5 \pm 9.6) \times 10^{-20}$

are several known techniques for improving optical clock performance beyond that which is demonstrated here—such as real-time blackbody-shift corrections (3), zero dead-time operation (30), and high-performance laser local oscillators (31)—that have led to lower instabilities as indicated in the purple line of Fig. 3. Separate measurements of the added instability due to noise in our optical-to-microwave transfer, shown in pink in Fig. 3, reach 5×10^{-17} at 1 s and 1×10^{-18} at 200 s (24). This indicates that our optical-to-microwave down-conversion can support the highest stability optical clocks yet demonstrated without degradation.

In addition to stability, we examined possible frequency offsets in the optical-to-electrical transfer that would degrade the accuracy of the resulting microwave signal (24). This is best analyzed by comparing the separation in the Yb clock frequencies as determined by the microwave measurement and as determined by the direct optical beat. Table 1 shows the results of our accuracy analysis and includes directly measured frequency differences without accounting for known systematic shift mechanisms in the clock systems (such as blackbody radiation-induced shifts). Both optical and microwave measurements yielded a fractional frequency offset near 5.9×10^{-17} , consistent with the known offset between the Yb clocks used in our experiments. More importantly, the difference in the offset from microwave and optical measurements, again represented as

a fractional offset, was only 2.5×10^{-20} . This is smaller than the statistical uncertainty of 9.6×10^{-20} of the point-by-point difference shown above in Fig. 2. Thus, any unintentional offsets resulting from the frequency transfer from the optical to the microwave domain are well below the $\sim 10^{-18}$ accuracy level of a state-of-the-art optical clock (3, 32).

Transferring the phase, the frequency stability, and the accuracy of optical clocks to the electronic domain has resulted in 100-fold improvement over the best microwave sources. With microwave signals having optical clock stability, one can envision a robust, phase-coherent system of ultrastable electronic signals capable of supporting future radar, communications, navigation, and basic science. Moreover, with residual instability of the optical-to-microwave link below that of the best optical clock demonstrations to date, further improvements to the absolute stability of microwaves can be expected.

REFERENCES AND NOTES

1. T. Rosenband *et al.*, *Science* **319**, 1808–1812 (2008).
2. R. M. Godun *et al.*, *Phys. Rev. Lett.* **113**, 210801 (2014).
3. W. F. McGrew *et al.*, *Nature* **564**, 87–90 (2018).
4. S. Kolkowitz *et al.*, *Phys. Rev. D* **94**, 124043 (2016).
5. A. D. Ludlow, M. M. Boyd, J. Ye, E. Peik, P. O. Schmidt, *Rev. Mod. Phys.* **87**, 637–701 (2015).
6. H. Katori, *Nat. Photonics* **5**, 203–210 (2011).
7. S. Weyers *et al.*, *Metrologia* **55**, 789–805 (2018).
8. F. Riehle, P. Gill, F. Arias, L. Robertsson, *Metrologia* **55**, 188–200 (2018).
9. J. Ye, J. L. Hall, S. A. Diddams, *Opt. Lett.* **25**, 1675–1677 (2000).

10. T. M. Fortier *et al.*, *Nat. Photonics* **5**, 425–429 (2011).
11. X. Xie *et al.*, *Nat. Photonics* **11**, 44–47 (2017).
12. C. Clivati *et al.*, *Sci. Rep.* **7**, 40992 (2017).
13. L. I. Gurvits, *Adv. Space Res.* **65**, 868–876 (2020).
14. D. J. Jones *et al.*, *Science* **288**, 635–640 (2000).
15. T. Udem, R. Holzwarth, T. W. Hänsch, *Nature* **416**, 233–237 (2002).
16. L. A. M. Johnson, P. Gill, H. S. Margolis, *Metrologia* **52**, 62–71 (2015).
17. N. Ohmae, N. Kuse, M. E. Fermann, H. Katori, *Appl. Phys. Express* **10**, 062503 (2017).
18. Y. Yao, Y. Jiang, H. Yu, Z. Bi, L. Ma, *Natl. Sci. Rev.* **3**, 463–469 (2016).
19. F. N. Baynes *et al.*, *Optica* **2**, 141–146 (2015).
20. W. Zhang *et al.*, *Appl. Phys. Lett.* **96**, 211105 (2010).
21. J. Davila-Rodriguez *et al.*, *Opt. Express* **26**, 30532–30545 (2018).
22. X. Xie *et al.*, *Optica* **1**, 429–435 (2014).
23. F. Quinlan *et al.*, *Nat. Photonics* **7**, 290–293 (2013).
24. See supplementary materials.
25. W. F. McGrew *et al.*, *Optica* **6**, 448–454 (2019).
26. J. A. Sherman, R. Jördens, *Rev. Sci. Instrum.* **87**, 054711 (2016).
27. A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, S. W. Director, Ed. (Communications and Signal Processing Series, McGraw-Hill, ed. 3, 1991).
28. L. C. Sinclair *et al.*, *Appl. Phys. Lett.* **109**, 151104 (2016).
29. J. G. Hartnett, N. R. Nand, C. Lu, *Appl. Phys. Lett.* **100**, 183501 (2012).
30. M. Schioppo *et al.*, *Nat. Photonics* **11**, 48–52 (2017).
31. E. Oelker *et al.*, *Nat. Photonics* **13**, 714–719 (2019).
32. S. M. Brewer *et al.*, *Phys. Rev. Lett.* **123**, 033201 (2019).
33. T. Nakamura, Data for “Coherent optical clock down-conversion for microwave frequencies with 10^{-18} instability.” NIST Public Data Repository (2020); doi:10.18434/M32206.

ACKNOWLEDGMENTS

We thank C. W. Oates for discussions about overall experiments and comments on the manuscript. We further thank J. C. Bergquist, D. Slichter, and L. C. Sinclair for their comments on this manuscript. **Funding:** This paper is supported by the National Institute of Standards and Technology and the DARPA PULSE program. **Author contributions:** T.N., J.D.-R., J.A.S., and F.Q. performed the microwave measurements. T.N., J.D.-R., H.L., T.M.F., S.A.D., and F.Q. developed the erbium fiber frequency combs. X.X. and J.C.C. designed and fabricated the high-speed photodetectors. W.M.F., X.Z., Y.S.H., D.N., K.B., and A.D.L. designed, constructed, and operated the Yb optical clocks. F.Q. supervised the work. All authors contributed to the final manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The data from the main text and supplementary materials are available from the NIST Public Data Repository (33). This is a contribution of the National Institute of Standards and Technology, not subject to U.S. copyright.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6493/889/suppl/DC1
Supplementary Text
Figs. S1 to S8
Table S1
References (34–49)
14 February 2020; accepted 17 April 2020
10.1126/science.abb2473

STRUCTURAL BIOLOGY

Ion transport and regulation in a synaptic vesicle glutamate transporter

Fei Li^{1,2}, Jacob Eriksen², Janet Finer-Moore¹, Roger Chang^{2,3*}, Phuong Nguyen¹, Alisa Bowen¹, Alexander Myasnikov^{1†}, Zanlin Yu¹, David Bulkley¹, Yifan Cheng^{1,4}, Robert H. Edwards^{2‡}, Robert M. Stroud^{1‡}

Synaptic vesicles accumulate neurotransmitters, enabling the quantal release by exocytosis that underlies synaptic transmission. Specific neurotransmitter transporters are responsible for this activity and therefore are essential for brain function. The vesicular glutamate transporters (VGLUTs) concentrate the principal excitatory neurotransmitter glutamate into synaptic vesicles, driven by membrane potential. However, the mechanism by which they do so remains poorly understood owing to a lack of structural information. We report the cryo-electron microscopy structure of rat VGLUT2 at 3.8-angstrom resolution and propose structure-based mechanisms for substrate recognition and allosteric activation by low pH and chloride. A potential permeation pathway for chloride intersects with the glutamate binding site. These results demonstrate how the activity of VGLUTs can be coordinated with large shifts in proton and chloride concentrations during the synaptic vesicle cycle to ensure normal synaptic transmission.

The storage of neurotransmitters inside synaptic vesicles enables their release by regulated exocytosis, conferring the vesicular (or quantal) release that mediates synaptic transmission (1). Synaptic vesicles take up classical neurotransmitters [monoamines, acetylcholine, γ -aminobutyric acid (GABA), and glutamate] from the cytosol, mediated by specific vesicular neurotransmitter transporters (VNTs) (2). A proton electrochemical gradient ($\Delta\mu_{H^+} = \Delta pH + \Delta\psi$) generated by the vacuolar adenosine triphosphatase (V-ATPase) across the synaptic vesicle membrane drives this uptake by all VNTs, but the VNTs vary in their dependence on the chemical gradient (ΔpH) and the membrane potential ($\Delta\psi$) component of $\Delta\mu_{H^+}$ (2). Vesicular glutamate transporters (VGLUTs) package the major excitatory neurotransmitter glutamate, driven predominantly by $\Delta\psi$ (3, 4). A $\Delta\psi$ of ~ 80 mV alone suffices to concentrate glutamate ~ 20 -fold to the observed luminal concentration of >100 mM, which enables the activation of postsynaptic receptors upon vesicle fusion and the release of concentrated neurotransmitter into the synaptic cleft. However, the mechanism by which these transporters function remains poorly understood in the absence of structural information.

As synaptic vesicles cycle at the nerve terminal, the rapidly changing ionic conditions also impose a series of challenges for the regulation of VGLUTs (2). The positive outside resting potential of the cell membrane resembles the synaptic vesicle membrane potential. Once vesicles have fused with the plasma membrane, VGLUTs become resident in the plasma membrane and could cause nonquantal release of glutamate because of the positive outside membrane potential. Upon reinternalization from the plasma membrane, vesicles trap extracellular solution with ~ 120 mM Cl^- and neutral pH, conditions that are unfavorable for glutamate filling. The high concentration of luminal glutamate required for synaptic transmission necessitates a corresponding displacement of the luminal Cl^- . To cope with these challenges, the VGLUTs exhibit complex interactions with H^+ and Cl^- (4–13). Additionally, excessive release of glutamate can produce excitotoxicity (14), and misregulation of the VGLUTs has been implicated in psychiatric and neurodegenerative diseases (15, 16). However, the mechanisms that underlie the regulation of VGLUTs have remained unidentified.

Mammals express three closely related VGLUT isoforms (75% sequence identity; fig. S1). The two major isoforms VGLUT1 and VGLUT2 exhibit complementary expression in, respectively, the cortex and diencephalon (17), and the loss of either impairs survival (18, 19). Because rat VGLUT2 is only 65 kDa, we determined its structure at 3.8-Å resolution by cryo-electron microscopy (cryo-EM) facilitated by an antigen-binding fragment (Fab) (Fig. 1A and figs. S2 and S3). Densities corresponding to lipids or detergents lie parallel to the VGLUT2 helices (fig. S4). The structure of VGLUT2 was determined de novo (fig. S5 and table S1) and adopts a canonical major facilitator superfamily (MFS)

fold (Fig. 1, B and C). Consistent with an MFS transporter that uses the alternating access mechanism, most transmembrane (TM) helices are distorted or kinked by proline and/or glycine (20). Reflecting its function in transporting a negatively charged substrate, the central cavity of VGLUT2 is positively charged (Fig. 1D).

Although VGLUT2 was captured in a luminal (outward) open apo state, comparison with other family members illuminates the basis for substrate specificity. The nine members of the SLC17 family in humans transport diverse organic anions, including glutamate (VGLUT1, VGLUT2, and VGLUT3), sialic acid (sialin), ATP [vesicular nucleotide transporter (VNUT)], and urates [sodium-phosphate transporters (NPTs)] (21) (Fig. 2A). Among their substrates, glutamate is the only one with two carboxyl groups. In VGLUT2, positively charged R88 orients toward the central binding site (Fig. 2B) and is conserved, which is consistent with a common role in anion recognition by the SLC17 family. The equivalent residue (R47) in the bacterial homolog D-galactonate transporter (DgoT) makes a salt bridge with the carboxyl group of its substrate D-galactonate (22) (Fig. 2C). R322 faces R88 from the opposite side of the binding site. Consistent with recognition of the second carboxyl in glutamate, R322 is conserved among the VGLUTs but not in other SLC17 family members. When glutamate is manually placed into the central cavity to mimic D-galactonate in DgoT, R88 and R322 can coordinate the two carboxyl groups (Fig. 2B). We tested the role of R88 and R322 in synaptic transmission by measuring miniature excitatory postsynaptic currents (mEPSCs) caused by the release of single synaptic vesicles from VGLUT1 and VGLUT2 double knockout hippocampal neurons rescued by wild-type (WT) and mutant VGLUT2. In contrast to the wild type, the R88A mutant drastically impairs synaptic transmission, whereas R322A eliminates release (Fig. 2D and fig. S6A). Both mutations also eliminate glutamate currents recorded from endosomes expressing VGLUT2 (13). We speculate that low levels of residual activity may enable R88 (but not R322) to fill synaptic vesicles undergoing spontaneous release, which have a long time to fill. The two arginines are well matched to the distance between substrate carboxyl groups, and, accordingly, aspartate is not transported by the VGLUTs (3, 5). In addition to R88 and R322, the substrate binding site is surrounded by aromatic and polar residues. Consistent with recognition of the carboxyl group common to SLC17 substrates, Y135 is also conserved (Fig. 2B).

Both R88 and R322 interact with clusters of charged and polar residues buried within the N- and C-domains through coulombic interactions (fig. S7, A to D). These networks

¹Department of Biochemistry and Biophysics, University of California San Francisco (UCSF) School of Medicine, San Francisco, CA, USA. ²Departments of Neurology and Physiology, UCSF School of Medicine, San Francisco, CA, USA. ³Graduate Program in Biomedical Sciences, UCSF, San Francisco, CA, USA. ⁴Howard Hughes Medical Institute, UCSF, San Francisco, CA, USA.

*Present address: Department of Neurology, University of Washington, School of Medicine, Seattle, WA, USA.

†Present address: Cryo-Electron Microscopy and Tomography Center, St. Jude Children's Research Hospital, Memphis, TN, USA.

‡Corresponding author. Email: stroud@msg.ucsf.edu (R.M.S.); robert.edwards@ucsf.edu (R.H.E.)

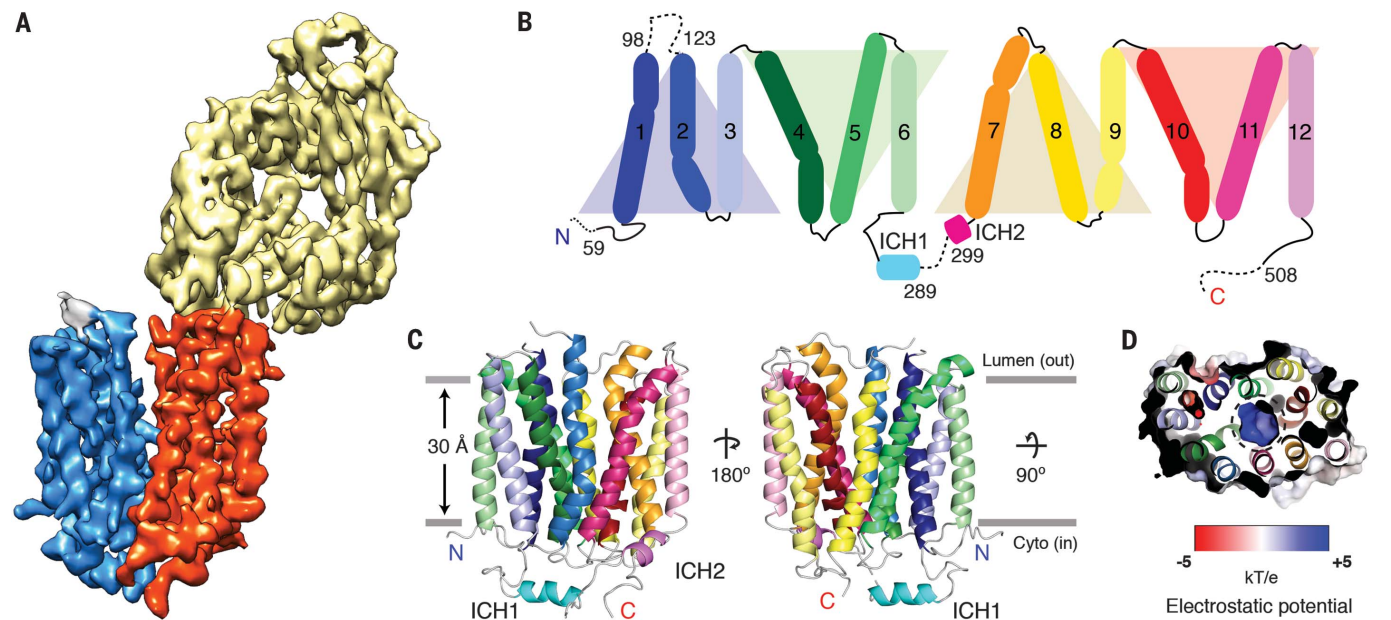
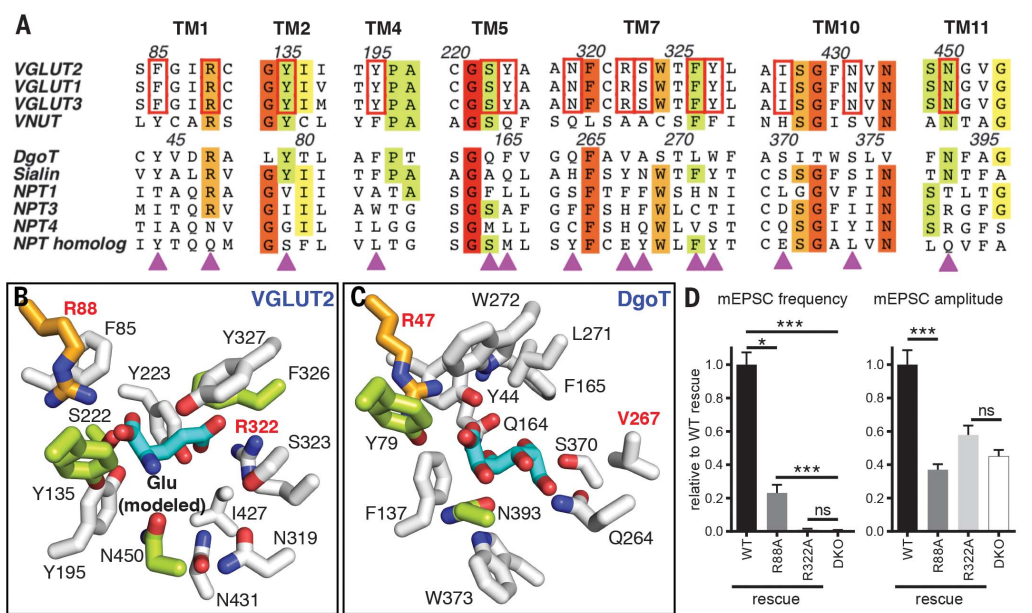


Fig. 1. Structure of VGLUT2. (A) Cryo-EM map of the VGLUT2-Fab complex. The two domains are colored blue (N-domain) and red (C-domain), and the Fab is colored yellow. (B) Schematic representation of the structural arrangement of VGLUT2. Three-helix bundles are related to each other by a twofold pseudosymmetry, and each bundle is colored using shades of the same color group. (C) Structure of VGLUT2. Helices are colored according to the

representation in (B), with connecting strands shown in gray. The VGLUT2 structure includes residues 59 to 508 except for the disordered loop 1 between TM1 and TM2 (residues 98 to 123) and 10 residues between ICH1 and ICH2 (residues 288 to 299). (D) Electrostatic surface of VGLUT2 shown at the plane parallel to the membrane through the central substrate binding site of the protein. The central cavity is indicated by a black dashed circle.

Fig. 2. R88 and R322 are critical for glutamate transport. (A) Alignment of human SLC17 family proteins, presented alongside the *Escherichia coli* homolog DgoT, with binding site residues indicated by magenta triangles. Residues are colored according to sequence conservation in descending order of red, orange, yellow, green, and no color. Putative substrate binding residues conserved in VGLUTs alone are highlighted by red boxes. (B) Structure of the substrate binding site in VGLUT2. Glutamate was manually placed into the binding site to mimic D-galactonate in DgoT. Both R88 and R322 are located at distances suitable for interacting with the carboxyl groups. (C) Structure of the substrate binding site in DgoT with substrate D-galactonate bound in the outward occluded conformation [PDB: 6E90 (22)]. Structures of VGLUT2 (B) and DgoT (C) are colored following the same pattern as (A), and substrates are colored cyan. (D) mEPSCs recorded from hippocampal neurons of VGLUT1 and VGLUT2 double knockout (DKO) mice rescued with WT, R88A, and R322A VGLUT2. Synaptic transmission is impaired by the R88A mutation and eliminated by the R322A mutation. mEPSC frequency (left) and amplitude (right) are normalized to those of VGLUT2-WT ($n = 10$ to 15 cells



per condition). Data indicate means and SEM. Statistical significance was determined by one-way analysis of variance (ANOVA) with Tukey's post hoc test. * $P < 0.05$; *** $P < 0.001$; ns, not significant. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

connect substrate binding by R88 and R322 with titratable regulatory sites within the transmembrane domains. The R88 cluster (Fig. 3, A to C) in the N-domain includes residues conserved throughout the SLC17 family. In contrast, the R322 cluster (Fig. 3, A and D) in the C-domain includes a histidine (H487)-glutamate (E396) pair 3.0 Å apart and C321, all of which are specifically conserved only in the VoGLUTs (fig. S1), which suggests a possible role in binding the second carboxyl group of glutamate.

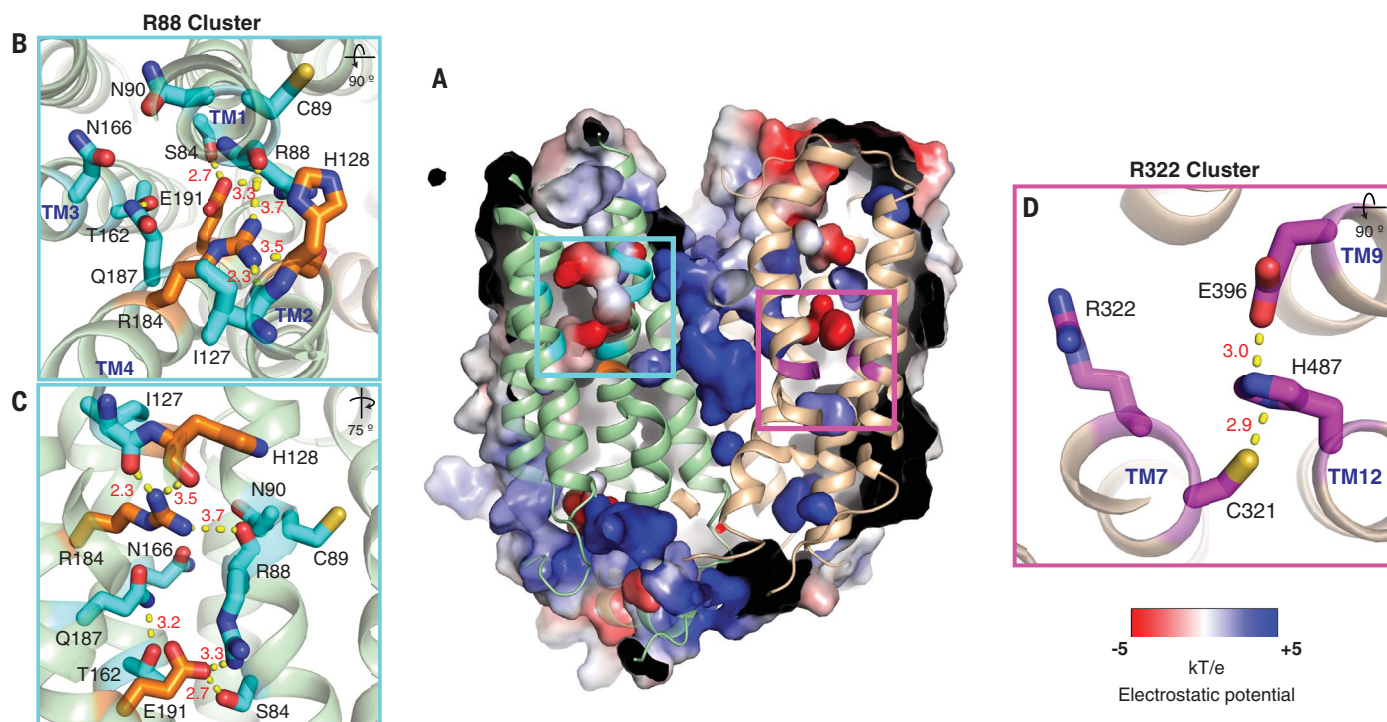


Fig. 3. Two functional clusters of charged and polar residues embedded within the transmembrane domains. (A) Electrostatic surface of VGLUT2. Two internal charged and polar cavities (referred to as R88 and R322 clusters) are colored according to the potential scale (bottom right). The N-domain is colored light green and the C-domain is tan. Cyan and magenta boxes in (A) match the insets in (B) (top view), (C) (side view), and (D) (top view), showing key residues and distances between polar groups (Å). (B and C) R88 cluster with proposed Cl^- (R184) and H^+ binding sites (E191 and H128) highlighted in orange. (D) R322 cluster.

The allosteric regulation of VGLUTs by H^+ provides a mechanism to prevent glutamate efflux to the synaptic cleft when VGLUTs translocate from the synaptic vesicle (pH ~5.6) to the presynaptic plasma membrane (pH ~7.4) during exocytosis. In hippocampal neurons, external high pH blocks the nonvesicular efflux of glutamate by VGLUT expressed on the plasma membrane (11, 23), which limits the potential for tonic excitation and excitotoxicity. Protons act allosterically to gate the VGLUT-associated Cl^- conductance (11) and, because of the similar regulation, we infer a similar allosteric mechanism for the activation of glutamate transport. In VGLUT2, E191 lies in the center of the N-domain within the buried, charged R88 cluster (Fig. 3, A to C). E191 is highly conserved in TM4 of the VGLUTs, and mutations of E191 reduce transport activity (24). The equivalent residue in DgoT (E133) is essential for H^+ symport (22) and is conserved in the H^+ symporter sialin (fig. S1). Consistent with a role in H^+ recognition, a glutamate is not present at this position in ATP or urate transporters, which are not H^+ driven (fig. S1). These factors and the environment of E191 support the idea of a role for E191 as a luminal H^+ sensor; protonation of this residue may liberate the adjacent R88 to interact with the substrate. In the case of DgoT and sialin, substrate transport is stoichiometrically coupled to the flux of H^+ . In contrast, VGLUT2 has lost the obliga-

tory coupling of H^+ to substrate, thereby minimizing the potential leakage of glutamate from the H^+ -rich synaptic vesicle (25) and allowing the uptake of glutamate against the synaptic vesicle H^+ gradient. Thus, E191 appears to have undergone a transition in role from H^+ coupling in DgoT and sialin to allosteric activation by H^+ in the VGLUTs. However, because E191 is buried, H128—partially exposed to the lumen and specific to VGLUTs—could act as the initial H^+ binding site from the lumen. It is also connected to E191 by a polar, water-filled tunnel with a minimum diameter of 1.8 Å (fig. S8A).

Located at the center of the R88 cluster, R184 in TM4 interacts with the backbone of I127 and H128 in TM2 as well as the backbone of R88 in TM1 (Fig. 3, B and C). R184 is proposed to confer allosteric activation by Cl^- because the neutralization of this residue eliminates the requirement for luminal Cl^- (13). R184 is conserved in all SLC17 proteins that generally depend on Cl^- (26). The positive charge on R184 interacts electrostatically with E191 and the entire R88 cluster (fig. S7, E and F). Hence, neutralization of the charge on R184 by Cl^- binding will have a large effect on the structure of the R88 cluster and will favor protonation of E191 (i.e., raise the pK_a , where K_a is the acid association constant). Similarly, protonation of E191 may favor Cl^- binding to R184.

The volume close to both R184 and H128—which are 5.2 Å apart—is large enough to accommodate a hydrated Cl^- ion (with a diameter of 3.6 Å). Tunnels with minimum diameters of 2.2 to 2.8 Å allow ready access for Cl^- to this site (fig. S8B). Chloride ions are most often coordinated by arginine, histidine, and serine (27), and hydrophobic surfaces are often seen surrounding tunnels, for example in Cl^- channels, or in excitatory amino acid transporters (EAATs). Thus, proximity of R184 to H128 may render the Cl^- binding subject to regulation by luminal acidic pH (11) and more than ~30 mM luminal Cl^- (13), so both are required to activate the VGLUTs. The current structure of VGLUT2 at pH 7.4, in an inhibited state, suggests that neutral pH may deprotonate H128 and so might inhibit Cl^- binding in the vicinity of H128 and R184. We do not see a Cl^- ion in the binding site under these conditions with 150 mM NaCl. However, Cl^- has a large negative form factor for electron scattering at low resolution (infinity to 16 Å) that may diminish any density. High-pass filtering to higher resolution did not reveal any additional density. Therefore, it is also reasonable that a weakly bound, hydrated chloride could be at this site in the structure. Consistent with their roles, mutations of R184 or H128 in VGLUT2 substantially reduce activity (24). The proximity of the H^+ and Cl^- binding sites explains the interdependence

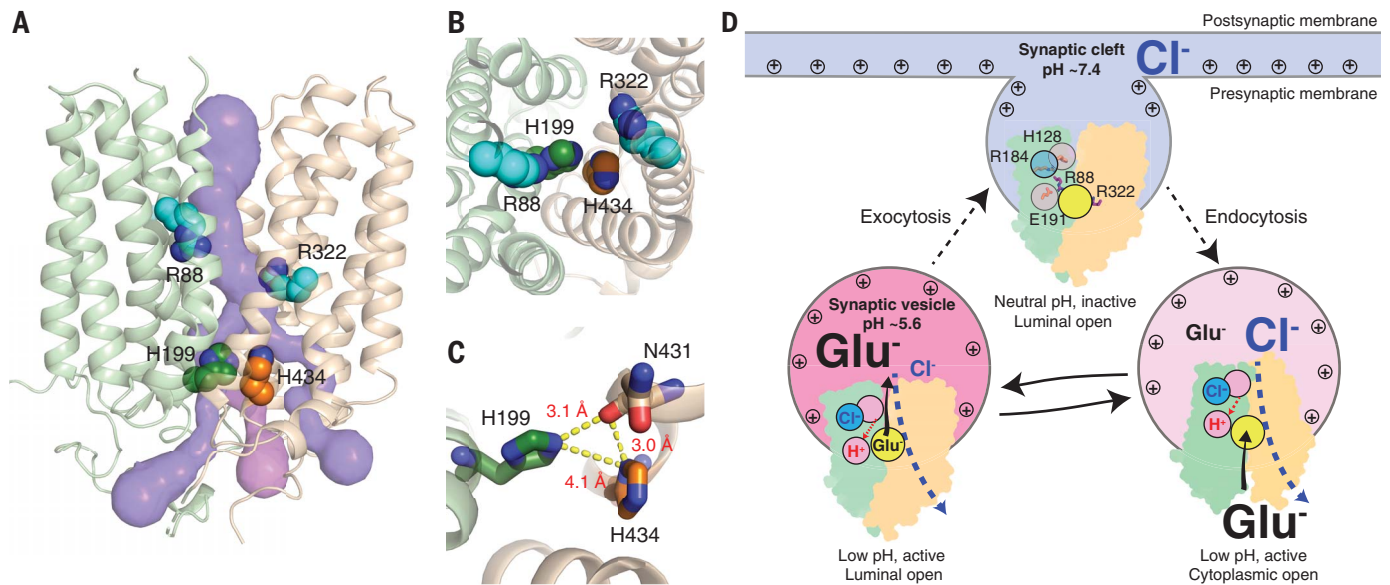


Fig. 4. Putative Cl^- channels and proposed transport mechanism. (A) Channels consistent with Cl^- conductance (surface). (B) Top view of the central channel. (C) Top view of the cytoplasmic two-His gate. Distances (Å) between polar groups are shown with dotted lines. (D) Proposed mechanism by which VGLUT integrates the dynamic change of ionic conditions during synaptic vesicle recycling to regulate its activity.

of pH and Cl^- that has been observed in the literature (4, 6, 11, 28).

It has been proposed that a VGLUT-associated Cl^- conductance enables the removal of Cl^- trapped during endocytosis to make room for glutamate (8, 11–13). The Cl^- conductance competes with glutamate transport (7), and mutation of the glutamate-binding R322 (as in R322A) abolishes the Cl^- conductance (11, 13) (fig. S6C). Together, these observations imply that the Cl^- channel intersects with the glutamate binding site. The R88A mutation also abolishes the Cl^- conductance of VGLUT2 (fig. S6B). The most commensurate channel, defined by a solvent-accessible surface of sufficient internal diameter to support a Cl^- ion, runs through the central cavity between substrate-binding R88 and R322 to a cytoplasmic gate formed by two histidine residues (H199 in the N-domain and H434 in the C-domain) that oppose each other (Fig. 4, A to C). This two-His gate is specific to the VGLUTs and sialin and is not present in other SLC17 family proteins. The two histidine residues are not hydrogen-bonded to each other; rather both H199 and H434 side chains donate hydrogen bonds to the backbone carbonyl of N431 (Fig. 4C). The hydrogen bond from H199 would contribute to the stabilization of the observed conformation, which is closed to the cytoplasmic side. Consistent with the role of H199 in forming and breaking this hydrogen bond that is critical for conformation change, a natural mutation of the equivalent residue in sialin (H183R) abolishes sialic acid transport and causes infantile sialic acid storage disease (29, 30). The channel narrows to 2.4-Å diameter at the two-His gate, which requires small rearrangements to

allow the conductance of Cl^- with a diameter of 3.6 Å. This parallels the Cl^- channel in TMEM16A, which has a similar diameter and characteristics and also narrows to 2.5 Å (31). With the two-His gate closed, the channel splits into two exits to the cytoplasmic side (Fig. 4A, purple). Both exit channels are too narrow to permit Cl^- permeation without rearrangement of the gate upon Cl^- entry. A third discontinuous channel (Fig. 4A, pink) leads from the two-His gate to the cytosol and is lined by residues conserved in the VGLUTs. This channel may also allow Cl^- to exit throughout the transport cycle.

On the basis of the structure presented in this study, we propose a transport mechanism and regulatory scheme for the VGLUTs (Fig. 4D). The electrochemical environment in the synaptic vesicles (low pH and >30 mM Cl^-) activates VGLUTs, whereas neutral pH at the plasma membrane inhibits VGLUTs, thus minimizing nonvesicular efflux. Under activating conditions, glutamate is transported through alternating access, bound by positive charges on R88 and R322 that stabilize the two negative charges of glutamate. This ensures that there is no insidious cotransport of H^+ on the substrate, against the proton gradient, into the vesicle, with its associated energy cost. This is in contrast to closely related H^+ -symporters DgoT and sialin. Upon substrate delivery by VGLUTs, the positive charge inside the vesicle favors the retention of the negatively charged substrate. Cl^- may contribute to neutralizing excess positive charge on VGLUT and promote recycling in the absence of substrate, thereby playing a role in the re-orientation from the luminal open structure. Thus, the overall transport cycle delivers sub-

strate to a lower energy state in the lumen and allows the concentration of glutamate to >100 mM. In the plasma membrane after exocytosis, neutral pH drives VGLUTs to an inactive, outward-open conformation where both the H^+ and Cl^- binding sites are empty (current structure). Transport would only resume after endocytosis, synaptic vesicle regeneration, and acidification, thereby integrating allosteric activation by H^+ at E191 mediated through H128, and by Cl^- at R184, with the synaptic vesicle cycle. A Cl^- channel intersecting with the glutamate binding site allows movement of Cl^- to balance electrostatic potential within—and osmotic forces across—the synaptic vesicle and to potentially determine an upper limit for vesicular glutamate concentration.

REFERENCES AND NOTES

1. B. Katz, *Science* **173**, 123–126 (1971).
2. R. H. Edwards, *Neuron* **55**, 835–858 (2007).
3. P. R. Maycox, T. Deckwerth, J. W. Hell, R. Jahn, *J. Biol. Chem.* **263**, 15423–15428 (1988).
4. J. S. Tabb, P. E. Kish, R. Van Dyke, T. Ueda, *J. Biol. Chem.* **267**, 15412–15418 (1992).
5. S. Naito, T. Ueda, *J. Neurochem.* **44**, 99–109 (1985).
6. H. Wolosker, D. O. de Souza, L. de Meis, *J. Biol. Chem.* **271**, 11726–11731 (1996).
7. E. E. Bellochio, R. J. Reimer, R. T. Freneau Jr., R. H. Edwards, *Science* **289**, 957–960 (2000).
8. S. Schenck, S. M. Wojcik, N. Brose, S. Takamori, *Nat. Neurosci.* **12**, 156–162 (2009).
9. G. Y. Goh et al., *Nat. Neurosci.* **14**, 1285–1292 (2011).
10. J. Preobraschenski, J. F. Zander, T. Suzuki, G. Ahnert-Hilger, R. Jahn, *Neuron* **84**, 1287–1301 (2014).
11. J. Eriksen et al., *Neuron* **90**, 768–780 (2016).
12. M. Martineau, R. E. Guzman, C. Fahlke, J. Klingauf, *Nat. Commun.* **8**, 2279 (2017).
13. R. Chang, J. Eriksen, R. H. Edwards, *eLife* **7**, e34896 (2018).
14. A. Lau, M. Tymianski, *Pflügers Arch.* **460**, 525–542 (2010).
15. A. Kashani, C. Betancur, B. Giros, E. Hirsch, S. El Mestikawy, *Neurobiol. Aging* **28**, 568–578 (2007).
16. A. Oni-Onisan, L. V. Kristiansen, V. Haroutunian, J. H. Meador-Woodruff, R. E. McCullumsmith, *Biol. Psychiatry* **63**, 766–775 (2008).

17. R. T. Freneau Jr. *et al.*, *Neuron* **31**, 247–260 (2001).
18. S. M. Wojcik *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 7158–7163 (2004).
19. A. Wallén-Mackenzie *et al.*, *J. Neurosci.* **26**, 12294–12307 (2006).
20. N. Yan, *Annu. Rev. Biophys.* **44**, 257–283 (2015).
21. R. J. Reimer, *Mol. Aspects Med.* **34**, 350–359 (2013).
22. J. B. Leano *et al.*, *PLoS Biol.* **17**, e3000260 (2019).
23. B. Mackenzie, A. C. Illing, M. E. Morris, H. Varoqui, J. D. Erickson, *Neurochem. Res.* **33**, 238–247 (2008).
24. N. Juge, Y. Yoshida, S. Yatsushiro, H. Omote, Y. Moriyama, *J. Biol. Chem.* **281**, 39499–39506 (2006).
25. C. Takami, K. Eguchi, T. Hori, T. Takahashi, *J. Physiol.* **595**, 1263–1271 (2017).
26. M. Hiasa, N. Togawa, Y. Moriyama, *Curr. Pharm. Des.* **20**, 2745–2749 (2014).
27. O. Carugo, *BMC Struct. Biol.* **14**, 19 (2014).
28. J. Hartinger, R. Jahn, *J. Biol. Chem.* **268**, 23122–23127 (1993).
29. P. Morin, C. Sagné, B. Gasnier, *EMBO J.* **23**, 4560–4570 (2004).
30. C. C. Wreden, M. Wlzl, R. J. Reimer, *J. Biol. Chem.* **280**, 1408–1416 (2005).
31. C. Paulino, V. Kalienkova, A. K. M. Lam, Y. Neldner, R. Dutzler, *Nature* **552**, 421–425 (2017).

ACKNOWLEDGMENTS

We thank M. Braunfeld and M. Harrington at the UCSF cryo-EM facility for their support in data acquisition and computation. We are also grateful for discussions with N. Poweleit, K. Verba, J.-P. Armache, M. Sun, A. Brilot, E. Green, and D. Asarnow regarding the cryo-EM. We also thank D. Cawley at the Vaccine and Gene Therapy Institute (Oregon Health and Science University) for generating the mAb and for advice on working with the antibodies; C. Craik, M. Bohn, and K. Basu for advice and help in characterizing and working with antibodies; and S. Schenck and R. Dutzler for sharing nanobodies against rVGLUT1 used during preliminary characterization and structural determinations efforts. **Funding:** This work is supported by R01NS089713 to R.M.S. and R.H.E. and R37MH50712 to R.H.E. F.L. is supported by postdoctoral fellowships from the American Heart Association (17POST33660928) and the National Institute of Mental Health (K99MH119591). The UCSF Electron Microscopy (EM) facility is supported by NIH grants S100DQ20054 and S100DQ21741. Y.C. is a Howard Hughes Medical Institute Investigator. **Author contributions:** F.L., R.M.S., and R.H.E. conceived the project. F.L., P.N., and A.B. expressed the protein. F.L. purified the protein; prepared all samples for EM; acquired cryo-EM data with the help of A.M., Z.Y., and D.B.; and

determined the structure of VGLUT2. Y.C. advised on cryo-EM experiments. J.E. and R.C. performed the functional experiments. F.L., J.F.-M., and R.M.S. analyzed the structure. F.L., J.F.-M., R.H.E., and R.M.S. wrote the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The atomic coordinates of rat VGLUT2 have been deposited in the Protein Data Bank with the accession code 6V4D. The corresponding map has been deposited in the Electron Microscopy Data Bank with the accession code EMD-21040.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6493/893/suppl/DC1
Materials and Methods
Figs. S1 to S10
Table S1
References (32–47)
MDAR Reproducibility Checklist

[View/request a protocol for this paper from Bio-protocol.](#)

27 January 2020; accepted 27 March 2020
10.1126/science.aba9202

CELL BIOLOGY

Supramolecular attack particles are autonomous killing entities released from cytotoxic T cells

Š. Bálint¹, S. Müller², R. Fischer³, B. M. Kessler³, M. Harkiolaki⁴, S. Valitutti^{2,5}, M. L. Dustin^{1,*}

Cytotoxic T lymphocytes (CTLs) kill infected and cancerous cells. We detected transfer of cytotoxic multiprotein complexes, called supramolecular attack particles (SMAPs), from CTLs to target cells. SMAPs were rapidly released from CTLs and were autonomously cytotoxic. Mass spectrometry, immunochemical analysis, and CRISPR editing identified a carboxyl-terminal fragment of thrombospondin-1 as an unexpected SMAP component that contributed to target killing. Direct stochastic optical reconstruction microscopy resolved a cytotoxic core surrounded by a thrombospondin-1 shell of ~120 nanometer diameter. Cryo-soft x-ray tomography analysis revealed that SMAPs had a carbon-dense shell and were stored in multicore granules. We propose that SMAPs are autonomous extracellular killing entities that deliver cytotoxic cargo targeted by the specificity of shell components.

Cytotoxic T lymphocytes (CTLs) exocytose soluble granzymes and perforin-1 (PRF1) from secretory lysosomes (SLs) into the interface between the CTL and target cell, the cytotoxic immunological synapse (IS) (1–5). PRF1 forms pores in the plasma membrane of target cells that mediate entry of granzymes, such as granzyme B (GZMB) into the target cell cytoplasm. Cytoplasmic GZMB initiates multiple pathways leading to target cell death (6–8). GZMB and PRF1 are stored inside SLs in condensates with serglycin (SRGN) (9, 10). There are reports that PRF1 and GZMB may be released in particles (11–13), but their nature has remained elusive. To address the mode of GZMB release, we designed an experiment to follow putative cytotoxic particles from cytomegalovirus phosphoprotein 65 (CMV pp65)-specific human CTL clones (14) to target cells bearing pp65-HLA-A2 complexes (HLA, human leukocyte antigen). The CTL clones were transfected with mRNA encoding GZMB-mCherry-SEpHluorin that concentrated in SLs. The SLs were also colabeled with Alexa Fluor 647 wheat germ agglutinin (WGA) (15). WGA does not interact with high mannose oligosaccharides of GZMB (16, 17), therefore cotransfer of GZMB and WGA to the target would implicate a multi-glycoprotein particle. The double-labeled CTLs were mixed with HLA-A2⁺ target cells with or without the pp65 peptide and subjected to time-lapse microscopy. Within minutes of mixing, the pp65 pulsed targets contained intense double-positive puncta, whereas unpulsed target cells lacked these signals after interaction with the

CTLs (Fig. 1A, fig. S1, and movies S1 to S3). We identified these multiprotein structures as supramolecular attack particles (SMAPs) and subjected them to further analysis.

We first investigated the kinetics of SMAP release. We incubated GZMB-mCherry-SEpHluorin transfected human CD8⁺ T cells on a supported lipid bilayer (SLB) coated with laterally mobile intercellular adhesion molecule-1 (ICAM-1) and anti-CD3ε (Fig. 1B and fig. S2) (18). Total internal reflection fluorescence microscopy (TIRFM) demonstrated that CTLs recruited acidic SLs displaying only mCherry fluorescence to the IS with activating SLB, which was rapidly followed (within 1 min) by the appearance of SEpHluorin puncta in the IS (Fig. 1B, fig. S2, and movie S4). Consistent with release of GZMB in a SMAP, the SEpHluorin signal persisted in the IS for 20 min rather than dispersing.

We next determined whether the SMAPs remained attached to the SLB after removal of the CTLs (Fig. 1C and movie S5). Untransfected CTLs were incubated on the activating SLB and either directly prepared for immunofluorescence detection of PRF1 and GZMB or removed before analysis (Fig. 1D). PRF1 and GZMB immunoreactivity was detected in the IS within 20 min, owing to the kinetics of antibody binding (figs. S3 and S4 and movies S6 to S9), and remained as discrete particles attached to the SLB after CTL removal (Fig. 1D). The SMAPs were stable without loss of PRF1 and GZMB for hours without fixation (fig. S5). We next tested SMAPs for their ability to kill target cells in a cytotoxicity assay in which dead cells release the cytoplasmic enzyme lactate dehydrogenase (LDH). Target cells were killed by SLB-immobilized SMAPs (Fig. 1E, black circles) after correction for spontaneous release of LDH by target cells (Fig. 1E, red circles). We also confirmed that SMAPs lacked LDH activity (Fig. 1E, blue triangles). Thus, SMAPs are stable after release from CTLs and can kill cells autonomously.

SMAPs captured on SLB were subjected to mass spectrometry (MS) analysis. We identified more than 285 proteins that were consistently present in SMAPs (Fig. 2, A and B). Of these, 82 proteins were unique to SMAPs on SLB with ICAM-1 and anti-CD3ε versus ICAM-1 alone, and 18 proteins were detected in most of the experiments (fig. S6). One peptide from PRF1 was detected in multiple experiments, and multiple GZMB peptides were identified in all experiments (fig. S6). We also identified a number of proteins involved in cell signaling (cytokines and chemokines) (fig. S6). The presence of PRF1 and GZMB in SMAPs was further confirmed by SDS-polyacrylamide gel electrophoresis (SDS-PAGE) and immunoblotting (fig. S7). Plasma membrane proteins such as the phosphatase CD45 and the degranulation marker lysosome-associated membrane glycoprotein 1 (LAMP-1, or CD107a) were not detected (fig. S7). This absence suggested minimal contamination with cellular membranes. Leukocyte function-associated antigen 1 (LFA-1) was confirmed by immunoblotting but not by immunofluorescence of SMAPs and thus may represent adhesion sites left on the SLB in parallel with SMAPs (19). Thrombospondin-1 (TSP1) stood out as a candidate because of its signature Ca²⁺-binding repeats (20, 21), which resonated with well-established Ca²⁺-dependent steps in CTL-mediated killing (22). Live imaging of the release of SMAPs on activating SLB showed that TSP1 and PRF1 are released together (fig. S8 and movie S10). In addition, TIRFM of SMAPs from CTLs transfected with full-length TSP1 with a C-terminal GFPspark revealed colocalization of the green fluorescent protein signal with GZMB and PRF1 antibody staining in the SMAPs (Fig. 2C and fig. S9), and anti-TSP1 antibody staining colocalized with mCherry and pHluorin signals from CTLs transfected with GZMB-mCherry-pHluorin (fig. S10). TSP1-GFPspark and GZMB-mCherry-SEpHluorin were colocalized within cytoplasmic compartments in cotransfected CTLs (fig. S11 and movie S11). This result suggested that SMAPs were preformed and stored in SLs. Enzyme-linked immunosorbent assays on soluble and SLB fractions from stimulation of primary CD8⁺CD57⁺ CTLs revealed similar levels of GZMB and PRF1 in both fractions, but the dependence on anti-CD3ε stimulation was higher for the SLB fraction (fig. S12). In contrast, TSP1 was almost exclusively in the SLB fraction and displayed significant dependence on anti-CD3ε stimulation (fig. S12). When we analyzed TSP1 by SDS-PAGE and immunoblotting, we found that CTLs and SMAPs contained not the full-length 145-kDa species stored in platelets but a C-terminal 60-kDa fragment under non-reducing and reducing conditions, which included the Ca²⁺-binding repeats (fig. S13) (23). CRISPR-Cas9-mediated knockout of TSP1 by 60% in CTLs reduced anti-CD3ε redirected

¹Kennedy Institute of Rheumatology, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, UK. ²Cancer Research Center of Toulouse, INSERM, Toulouse, France. ³Discovery Proteomics Facility, Target Discovery Institute, Nuffield Department of Medicine, University of Oxford, Oxford, UK. ⁴Diamond Light Source, Harwell Science and Innovation Campus, Chilton, Didcot, UK. ⁵Department of Pathology, Institut Universitaire du Cancer-Oncopole, Toulouse, France.

*Corresponding author. Email: michael.dustin@kennedy.ox.ac.uk

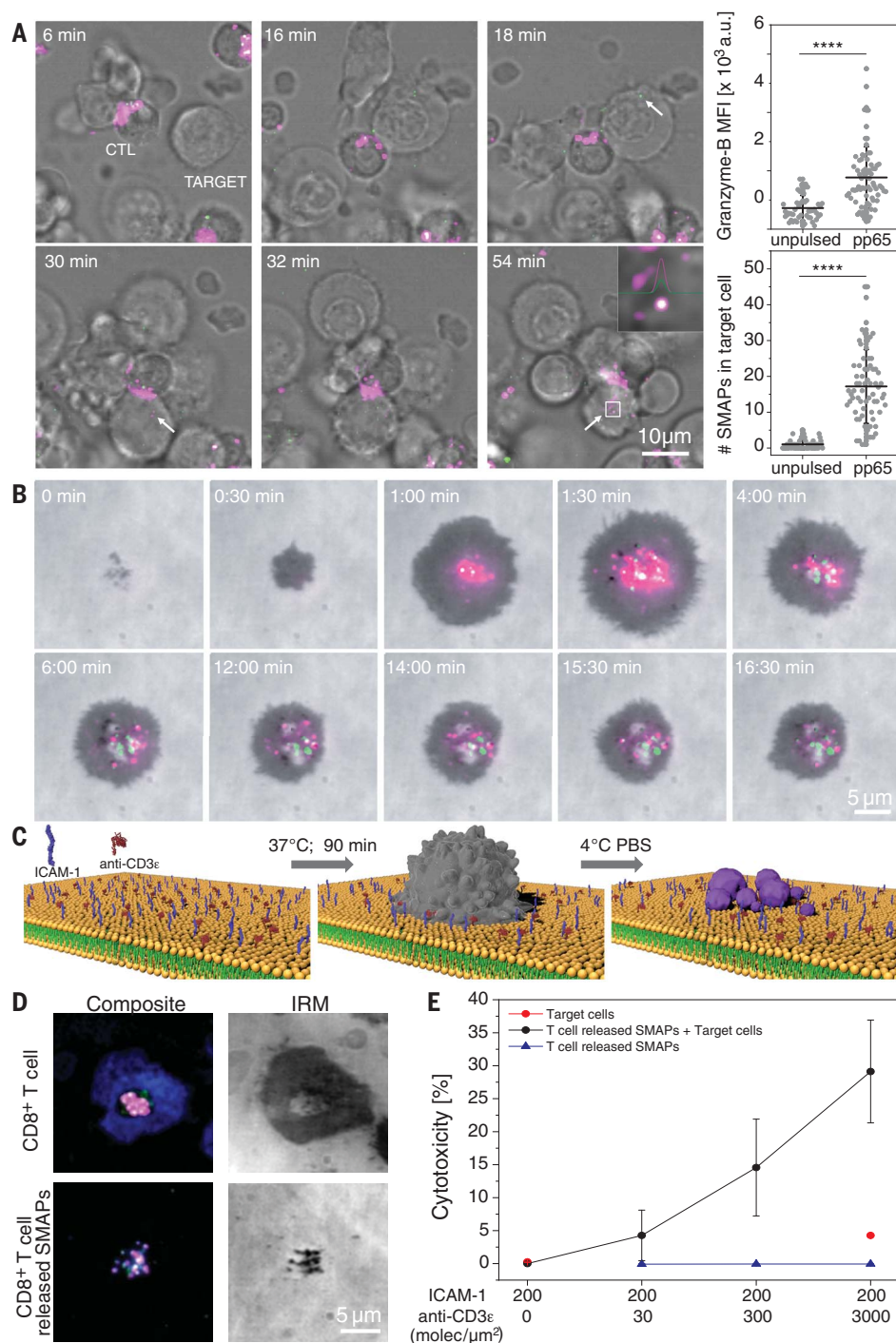


Fig. 1. SMAPs were released at the IS and displayed autonomous cytotoxicity. (A) Time-lapse confocal images depicting the transfer of SMAPs labeled with Gzmb-mCherry⁺ (green) and WGA (magenta) from an antigen-specific CTL clone into pp65-pulsed JY target cells ("TARGET"). Arrows and inset indicate the presence of SMAPs inside the target. Scale bar, 10 μ m. Quantification of GZMB mean fluorescence intensity (MFI) and number of double-positive particles inside the target cell in CTL conjugates with unpulsed or pulsed target cells. Each dot represents one target cell (<50 cells). Horizontal lines and error bars represent mean \pm SD from two independent experiments. **** P < 0.0001. (B) Live-cell imaging of SMAPs release by CD8⁺ T cells transfected with Gzmb-mCherry-SepHluorin (magenta and green) on activating SLB. IRM, interference reflection microscopy. Scale bar, 5 μ m. (C) Schematic of the working model for capturing SMAPs released by activated CD8⁺ T cells. CD8⁺ T cells (gray) were incubated on SLB presenting activating ligands for the indicated time. Cells were removed with cold phosphate-buffered saline (PBS) leaving the released SMAPs (purple) on the SLB. Elements are not drawn to scale. (D) TIRF images of CD8⁺ T cells incubated on activating SLB in the presence of anti-PRF1 (green) and anti-GZMB (magenta) antibodies (top panels). After cell removal, PRF1⁺ and GZMB⁺ SMAPs remained on the SLB (bottom panels). The formation of a mature IS is indicated by an ICAM-1 ring (blue). Scale bar, 5 μ m. (E) Target cell cytotoxicity induced by density-dependent release of SMAPs captured on SLB measured by LDH release assay. Data points and error bars represent mean \pm SEM from three independent experiments.

killing of K562 cells by 30% ($n = 5$ donors, $P < 0.001$), whereas knockout of another similarly enriched protein, galectin-1, by 90% had no effect on killing of K562 cells (Fig. 2, D and E). Although TSP1 is associated with T cell adhesion to extracellular matrix (24), TSP1 knockout did not alter T cell adhesion to activating SLB, but it did reduce the signals for TSP1, PRF1, and GZMB in SMAPs (fig. S14). These results suggested that the C-terminal domain of TSP1 was a component of SMAPs and was important in CTL-mediated killing.

We next investigated the organization of molecules within SMAPs at 20-nm resolution by direct stochastic optical reconstruction microscopy (dSTORM). SMAPs were detected with WGA in clusters of 27 ± 12 SMAPs per IS (Fig. 3A). On closer inspection, WGA staining appeared as a dense ring in the two-dimensional (2D) projections, which indicated a spherical shell with an average diameter of 120 ± 43 nm (Fig. 3A). Many supramolecular assemblies use phospholipid bilayers as a scaffold, so we sought to determine whether SMAPs stain

with the lipophilic membrane dye DiD, which brightly stains extracellular vesicles and lipoproteins. DiD did not stain SMAPs (fig. S15), which is consistent with the paucity of membrane proteins detected by MS (data S1). The WGA staining pattern was most consistent with a shell of glycoproteins (16) rather than a phospholipid-based membrane surrounding SMAPs. The location of TSP1 in SMAPs was investigated by multicolor dSTORM. Notably, TSP1 colocalized with WGA ($59 \pm 3\%$) and similarly highlighted the shape of the SMAPs (Fig.

3B and fig. S16). Thus, SMAPs from CTLs have a glycoprotein shell that includes TSP1.

To further investigate the structure of SMAPs, we used cryo-soft x-ray tomography (CSXT), a nondestructive 3D method based on the preferential absorption of x-rays by carbon-rich cellular structures in unstained, vitrified

specimens with a resolution of 40 nm (25, 26). For this imaging method, CTLs were incubated on electron microscopy (EM) grids coated with ICAM-1 and anti-CD3 ϵ . After incubation, samples were plunge-frozen with the T cells in place or removed to leave only the SMAPs. Released SMAPs captured on the grid after

cell removal (Fig. 3C and movie S12) were readily resolved and had an average diameter of 111 ± 36 nm (fig. S17). The slightly larger SMAP size determined by dSTORM reflected the contribution of ~ 9 nm due to the 2.45 nm hydrodynamic radius of WGA. The carbon-dense shell observed in CSXT was consistent with the TSP1-WGA shell observed by dSTORM. The CSXT analysis further emphasized intracellular multicore granules in the CTLs that appeared to be tightly packed with SMAPs, where the lower density cores were resolved (movie S13). These multicore granules were associated with the basal surface of CTLs near activating grids (Fig. 3D and movie S14), as expected (3).

We next performed three-color dSTORM to determine the location of cytotoxic proteins within SMAPs. The TSP1-WGA shell enclosed partly overlapping PRF1 $^{+}$ and GZMB $^{+}$ areas across the 2D projection (Fig. 4, A and B). We also detected SRGN in the core of SMAPs (fig. S18). Given the apparent density of material in the shell and the stability of SMAPs, it was notable that 150-kDa antibodies had access to components in the core. This is an unexpected property that will require further investigation. SMAPs containing PRF1 and/or GZMB were bigger and more abundant than WGA $^{+}$ particles devoid of cytotoxic proteins (Fig. 4, C and D). Primary CD8 $^{+}$ CD57 $^{+}$ CTLs and natural killer (NK) cells from peripheral blood also released SMAPs with PRF1, GZMB, and TSP1 (fig. S19). These results completed our picture of SMAPs: autonomously cytotoxic particles with an average diameter of ~ 111 nm with a dense shell including TSP1 and a core of PRF1, GZMB, and SRGN with unexpected accessibility to antibodies.

CTLs can also use the ligand for the death receptor Fas (FasL) to kill targets expressing Fas (27). We only detected FasL in the CTL IS when Fas glycoprotein was incorporated in the SLB with ICAM-1 and anti-CD3 ϵ (fig. S20). In these cases, FasL distribution in the IS was in puncta distinct from PRF1 and GZMB. The related protein CD40L is released in a CD40-dependent manner in helper T cell IS within synaptic ectosomes (28, 29). Synaptic ectosomes are a type of extracellular vesicle similar to exosomes but generated by budding from the plasma membrane of the T cell in the IS (29, 30). These results suggested that there were two types of cytotoxic particles released by CTLs in contact with Fas-expressing targets—FasL clusters in vesicles and SMAPs.

Our working model for SMAP function is that SMAPs act as autonomous killing entities with innate targeting through TSP1 and potentially other shell components. While SMAPs transferred through the IS may only affect one target, CTLs can kill without an IS, using a process involving rapid contacts (14, 31). The ability of SMAPs to autonomously kill target

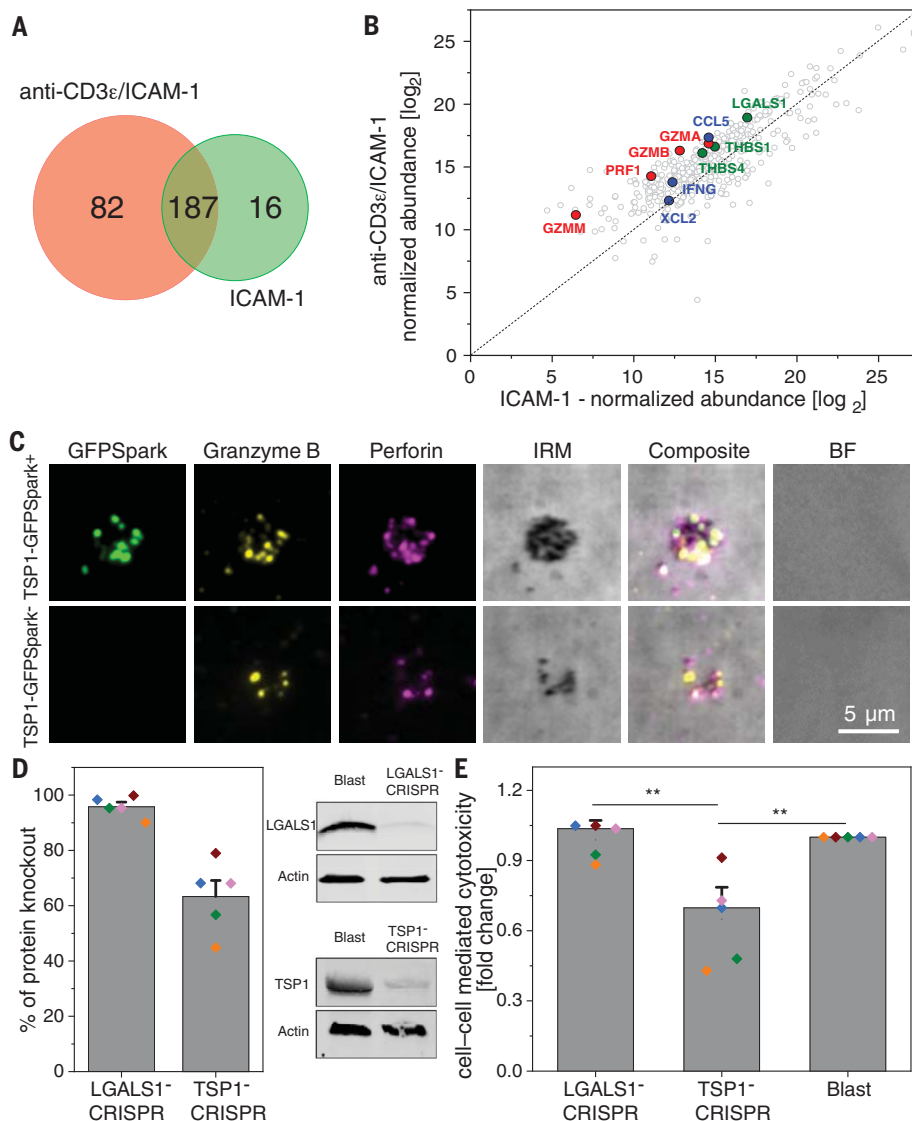


Fig. 2. TSP1 was a major constituent of SMAPs and contributed to CTL killing of targets. (A) Two-set Venn diagram showing the number of individual and common proteins identified by MS analysis of material released by CD8 $^{+}$ T cells incubated on nonactivating (ICAM-1) or activating (ICAM-1 + anti-CD3 ϵ) SLB. Representative of three independent experiments with eight donors. (B) Normalized abundance of the 285 proteins identified by MS in each condition. Cytotoxic proteins are highlighted in red, chemokines and cytokines in blue, and adhesion proteins in green. (C) TIRFM images of SMAPs released from CD8 $^{+}$ T cells transfected with TSP1-GFPspark (green, top row) or nontransfected cells (bottom row). Released SMAPs were further stained with anti-GZMB (yellow) and anti-PRF1 (magenta) antibodies. BF, bright-field microscopy. Scale bar, 5 μ m. (D) Percentage of galectin-1 and TSP1 knockout in CD8 $^{+}$ T cells by CRISPR-Cas9 genome editing measured from immunoblotting analysis (left). Each colored dot represents one donor. Bars represent mean \pm SEM. Representative immunoblot for galectin-1 (Lgals1) and TSP1 in Lgals1 $^{-}$ and TSP1-edited CD8 $^{+}$ T cells, respectively (right). CD8 $^{+}$ T cells (Blast) were analyzed in parallel as a control. (E) Target cell cytotoxicity mediated by galectin-1 (Lgals1-CRISPR) or TSP1 (TSP1-CRISPR) gene edited CD8 $^{+}$ T cells measured by LDH release assay. T cell blasts were used as a control. Bars represent mean \pm SEM. $^{**}P < 0.01$. Donors are the same as in (D).

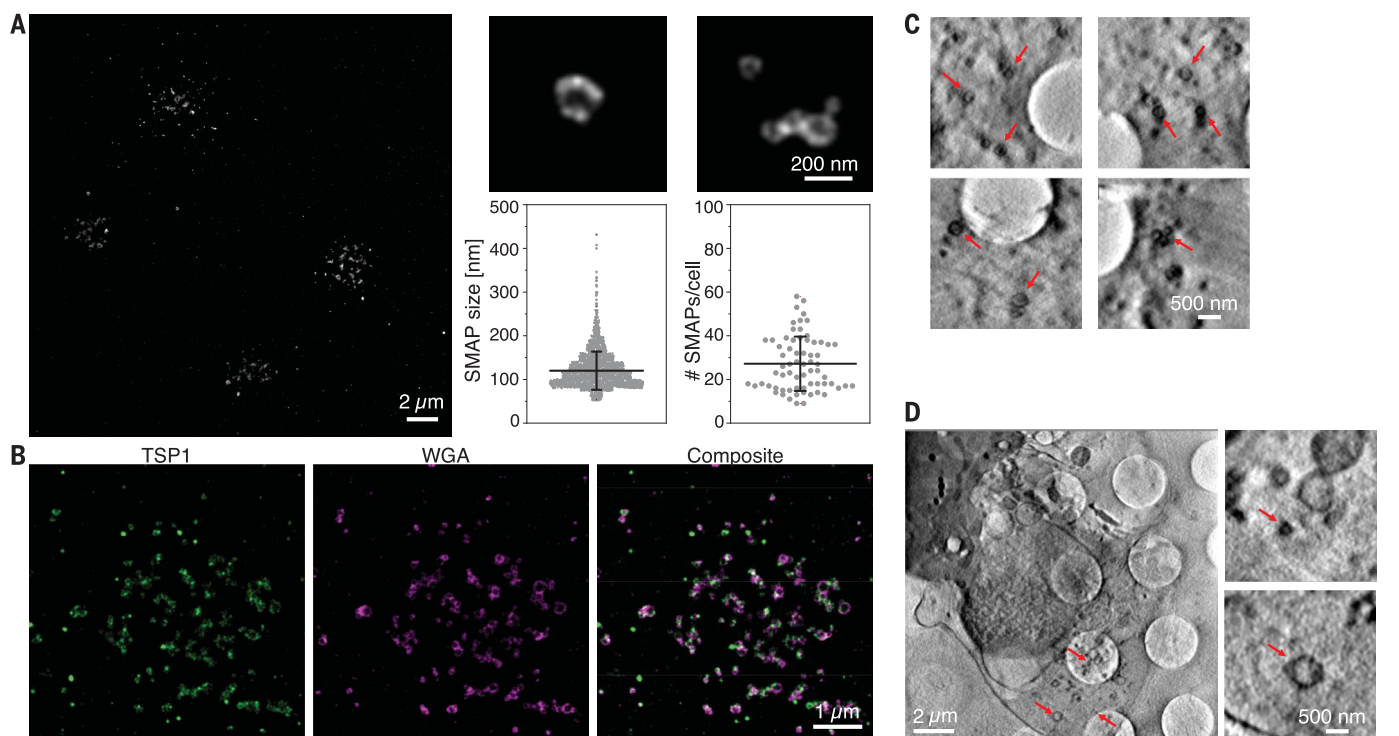


Fig. 3. SMAPs shell was rich in glycoproteins, TSP1, and organic material.

(A) dSTORM image of SMAPs released on activating SLB by multiple cells (left; scale bar, 2 μ m) and two examples of individual SMAPs (top right; scale bar, 200 nm), showing their heterogeneity in size. SMAPs were labeled with WGA. Quantification of SMAP size and number released per cell (bottom right; $n > 1800$ and $n = 67$, respectively). Horizontal lines and error bars represent

mean \pm SD from five donors. (B) dSTORM images of SMAPs (labeled with WGA, magenta) positive for TSP1 (green) released on activating SLB. Scale bar, 1 μ m. (C) Multiple CSXT examples of released SMAPs after cell removal. Scale bar, 500 nm. (D) CSXT of CD8⁺ T cells interacting with carbon-coated EM grids [note grid holes in (C) and (D)] containing ICAM-1 and anti-CD3 ϵ . Scale bars, 2 μ m (left) and 500 nm (right). Red arrows in (C) and (D) indicate SMAPs.

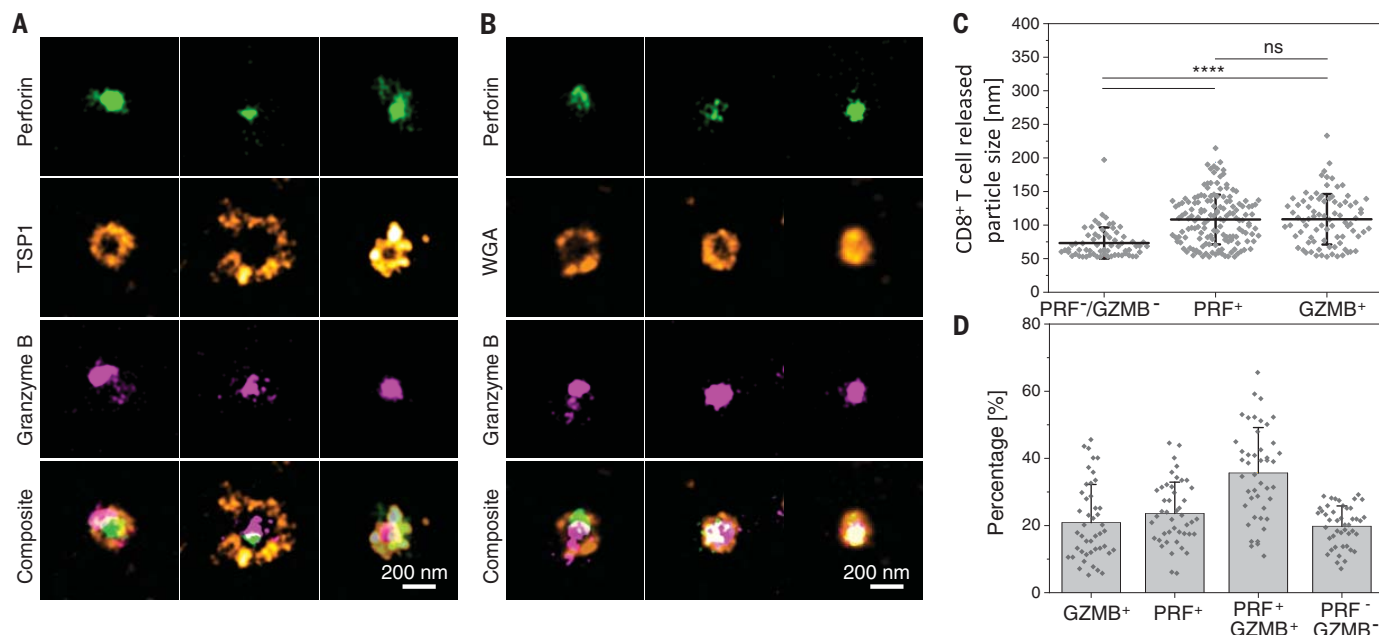


Fig. 4. SMAPs had a TSP1 shell and a core of cytotoxic proteins. (A and B) dSTORM images of individual SMAPs positive for PRF1 (green), GZMB (magenta), and TSP1 [(A), orange] or stained with WGA [(B), orange]. Scale bars, 200 nm. (C) Quantification of the size of cytotoxic particles on the basis of their protein composition ($n = 64$ for PRF1⁻ and GZMB⁻ cytotoxic particles, $n = 149$ and $n = 83$ for PRF1⁺ and GZMB⁺ cytotoxic particles, respectively). **** $P < 0.0001$. n.s., not significant. (D) Quantification of the percentage of particles positive and negative for PRF1 or GZMB. (C and D) Horizontal lines and bars and error bars represent mean \pm SD from five donors.

cells may become important in situations of transient and multiple interactions, where delivery might be less precise. SMAPs may have other modes of action, potentially including chemoattraction through C-C motif chemokine ligand 5 (CCL5, or RANTES) and immune modulation through interferon- γ . The TSP1 C terminus contains the binding site for the ubiquitous “don’t eat me” signal CD47 (32). SMAPs may thus partner with macrophages to ensure that any cell that cannot be killed by SMAPs is culled by phagocytosis (33).

REFERENCES AND NOTES

1. D. Masson, J. Tschopp, *Cell* **49**, 679–685 (1987).
2. J. Tschopp, S. Schäfer, D. Masson, M. C. Peitsch, C. Heusser, *Nature* **337**, 272–274 (1989).
3. J. C. Stinchcombe, G. Bossi, S. Booth, G. M. Griffiths, *Immunity* **15**, 751–761 (2001).
4. G. Ménasché, J. Feldmann, A. Fischer, G. de Saint Basile, *Immunol. Rev.* **203**, 165–179 (2005).
5. M. Kalos *et al.*, *Sci. Transl. Med.* **3**, 95ra73 (2011).
6. C. J. Froelich *et al.*, *J. Biol. Chem.* **271**, 29073–29079 (1996).
7. D. Keefe *et al.*, *Immunity* **23**, 249–262 (2005).
8. R. H. Law *et al.*, *Nature* **468**, 447–451 (2010).
9. S. S. Metkar *et al.*, *Immunity* **16**, 417–428 (2002).
10. J. A. Martina *et al.*, *Cell. Immunol.* **271**, 267–279 (2011).
11. P. J. Peters *et al.*, *Eur. J. Immunol.* **19**, 1469–1475 (1989).
12. P. J. Peters *et al.*, *J. Exp. Med.* **173**, 1099–1109 (1991).
13. L. Lugini *et al.*, *J. Immunol.* **189**, 2833–2842 (2012).
14. F. Bertrand *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 6073–6078

- (2013).
15. W. T. Kim *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 17143–17148 (2002).
16. R. Kornfeld, S. Kornfeld, *Annu. Rev. Biochem.* **54**, 631–664 (1985).
17. G. M. Griffiths, S. Isaza, *J. Cell Biol.* **120**, 885–896 (1993).
18. K. Somersalo *et al.*, *J. Clin. Invest.* **113**, 49–57 (2004).
19. J. C. Kuo, X. Han, C. T. Hsiao, J. R. Yates 3rd, C. M. Waterman, *Nat. Cell Biol.* **13**, 383–393 (2011).
20. M. Kvensakul, J. C. Adams, E. Hohenester, *EMBO J.* **23**, 1223–1233 (2004).
21. T. M. Misenheimer, D. F. Mosher, *J. Biol. Chem.* **270**, 1729–1733 (1995).
22. E. R. Podack, P. J. Konigsberg, *J. Exp. Med.* **160**, 695–710 (1984).
23. D. S. Annis, J. E. Murphy-Ullrich, D. F. Mosher, *J. Thromb. Haemost.* **4**, 459–468 (2006).
24. S. S. Li *et al.*, *Eur. J. Immunol.* **32**, 1069–1079 (2002).
25. G. Schneider, *Ultramicroscopy* **75**, 85–104 (1998).
26. M. Harkiolaki *et al.*, *Emerg. Top. Life Sci.* **2**, 81–92 (2018).
27. G. Bossi, G. M. Griffiths, *Nat. Med.* **5**, 90–96 (1999).
28. K. Choudhuri *et al.*, *Nature* **507**, 118–123 (2014).
29. D. G. Saliba *et al.*, *eLife* **8**, e47528 (2019).
30. E. Cocucci, J. Meldolesi, *Trends Cell Biol.* **25**, 364–372 (2015).
31. S. Halle *et al.*, *Immunity* **44**, 233–245 (2016).
32. V. Mateo *et al.*, *Nat. Med.* **5**, 1277–1284 (1999).
33. P. A. Oldenborg, H. D. Gresham, F. P. Lindberg, *J. Exp. Med.* **193**, 855–862 (2001).

ACKNOWLEDGMENTS

We thank E. Kurz, S. Valvo, L. Chen, H. Rada, M. C. Spink, M. P. Puissegur, M. Mixon, J. Love, and A. Sessions for important contributions to this work. We thank B. Alarcon, C. Baldari, J. Rettig,

E. Abu Shah, G. Griffiths, and F. Bravo Lopes for advice. **Funding:** Supported by the ERC AdG 670930 (M.L.D.); the Wellcome Trust 100262 (M.L.D.); the Kennedy Trust (to M.L.D. and B.M.K.); Versus Arthritis RACE 20298 (M.L.D.); the Laboratoire d’Excellence Toulouse Cancer (TOUCAN) under contract ANR11-LABX and Ligue Nationale contre le Cancer (Equipe labellisée 2018) (S.V.); the Institute for Protein Innovation; and STFC and Wellcome for support of Diamond Light Source Ltd. **Author contributions:** S.B. and M.L.D. conceived of the study and designed experiments. S.B. performed experiments and analyzed data. S.M. and S.V. performed CTL clone experiments. R.F. and B.M.K. acquired and analyzed MS data. M.H. acquired and analyzed CSXT data. S.B. and M.L.D. wrote the manuscript. **Competing interests:** S.B. and M.L.D. of the University of Oxford have filed UK Patent Application No. 2005892.1 on the isolation, engineering, and use of proteinaceous particles (SMAPs) in treatments of cancer and other disorders. **Data and materials availability:** All data are available in the manuscript or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/368/6493/897/suppl/DC1
Materials and Methods
Figs. S1 to S20
References (34–42)
MDAR Reproducibility Checklist
Movies S1 to S14
Data S1

[View/request a protocol for this paper from Bio-protocol.](#)

29 July 2019; resubmitted 28 January 2020
Accepted 9 April 2020
Published online 7 May 2020
10.1126/science.aay9207



THE FIRST SHANGHAI UNIVERSITY LEAGUE'S FORUM FOR INTERNATIONAL YOUNG SCHOLARS IS FORTHCOMING

Shanghai, a city of innovation, rich culture and ecology, is now developing into an extraordinary modern metropolis, and becoming a global center of economy, finance, trade, shipping, science and technology.

For the purpose of paving the way for young scholars to learn about the city and work in high schools, the First Shanghai University League's Forum for International Young Scholars (hereafter referred to as SULF) will be officially launched at 20:20 on May 30, 2020 (Beijing time) . We are looking forward to your online participation!

SULF is an international academic exchange activity for young scholars. It is hosted by the Shanghai Educational Human Resource Exchange and Service Center and Shanghai University Human Resource League, and organized by universities in Shanghai. SULF provides a all-round network-based platform and a professional service team composed of academic elites to build a bridge of cooperation and exchange between young scholars and universities in Shanghai. The forum aims to create a “carnival” for international young scholars. It establishes a platform for scholars to present themselves, seek cooperation, and explore Shanghai culture through activities such as lectures, academic exchanges, inter-school activities, and cultural exchange.

SULF is divided into 20 special sessions, covering a wide range of academic disciplines including mechanical engineering, civil engineering, textile science and engineering, chemical engineering, aquaculture, education, medicine, urban and rural planning, materials science, electronic science and technology, artificial intelligence, biomedicine, optical engineering, business management, law, and drama and film studies. It is hosted by 14 universities in Shanghai, including Fudan University, Shanghai Jiao Tong University, Tongji University, Shanghai University, and ShanghaiTech University and so on. Nobel Prize winners, top experts and academicians from various countries will be invited to make keynote speeches on the forum, and young scholars will be invited to interact and share their ideas.

Website: <https://shehr.shec.edu.cn/>

ADVERTISEMENT

SHANGHAI UNIVERSITY LEAGUE'S FORUM FOR INTERNATIONAL YOUNG SCHOLARS



Scan for more information

SCHOOL	MAIN FORUM	SCHOOL	MAIN FORUM
	Shanghai University League's Forum for International Young Scholars — Elec.Forum @FUDAN		Shanghai University League's Forum for International Young Scholars — GREEN ENG@SHANG HAI 2020
	Shanghai University League's Forum for International Young Scholars — Mechanical Engineering Forum @Shanghai Jiao Tong University		Shanghai University League's Forum for International Young Scholars — Textile Sci & Tech Forum @DHU
	Shanghai University League's Forum for International Young Scholars — MedForum @SJTUSM		Shanghai University League's Forum for International Young Scholars — Optical Engineering Forum @USST
	Shanghai University League's Forum for International Young Scholars — Autonomous Intelligent Systems Forum @Tongji		Shanghai University League's Forum for International Young Scholars — Forum of Theatre, Film and TV @Shanghai Theatre Academy
	Shanghai University League's Forum for International Young Scholars — Environment and Ecology Forum		Shanghai University League's Forum for International Young Scholars — Law Forum @ECUPL
	Shanghai University League's Forum for International Young Scholars — Stem Cell & Translational Medicine Forum @Tongji		Shanghai University League's Forum for International Young Scholars — Fishery Forum @Shanghai Ocean University
	Shanghai University League's Forum for International Young Scholars — Civil Engineering Forum @Tongji		Shanghai University League's Forum for International Young Scholars — Materials Frontier @SHU
	Shanghai University League's Forum for International Young Scholars — Architecture, Urban and Rural Planning & Landscape Architecture Forum @Tongji		Shanghai University League's Forum for International Young Scholars — forum for business studies @SHU
	Shanghai University League's Forum for International Young Scholars — Geo.Forum @East China Normal University		Shanghai University League's Forum for International Young Scholars — Elite Forum for Chinese Studies @Shanghai Normal University
	Shanghai University League's Forum for International Young Scholars — Edu.Forum @East China Normal University		Shanghai University League's Forum for International Young Scholars — Bio Forum @ShanghaiTech

ScienceCareers

FROM THE JOURNAL SCIENCE  AAAS

Confused about your next career move?



**Download Free Career
Advice Booklets!**

ScienceCareers.org/booklets





myIDP:
A career plan customized for you, by you.

For your career in science, there's only one **Science**

Features in myIDP include:

- Exercises to help you examine your skills, interests, and values.
- A list of 20 scientific career paths with a prediction of which ones best fit your skills and interests.



Visit the website and start planning today!
myIDP.sciencecareers.org

ScienceCareers In partnership with:







European Environment Agency



Call for Expressions of Interest for experts to be appointed members of the Scientific Committee of the European Environment Agency (EEA)

EEA/SC/2020/1-7

Scope of the call

The EEA organises this call to appoint approximately 10 Scientific Committee members with expertise in one or more of the following priority areas as identified by the EEA Management Board:

Data and information technology, Science Communication, Environment and Health/Chemicals, Circular economy and resource use, Biodiversity and ecosystems, Climate change, Sustainability and foresight.

Interested candidates are invited to look at the full vacancy notice published on the EEA website:
<http://www.eea.europa.eu/about-us/governance/scientific-committee/call-for-expression-of-interest>

Candidates must be nationals of one of the 32 EEA member countries.

Deadline for applications: Friday 3rd July 2020, 12.00 CET

Where Science Gets Social.

AAAS.ORG/COMMUNITY





Member COMMUNITY

AAAS

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

Jefferson Science Fellowship Program

The National Academies of Sciences, Engineering, and Medicine is pleased to announce a call for applications for the 2021 Jefferson Science Fellows (JSF) program. Established by the Secretary of State in 2003, this fellowship program engages the American academic science, technology, engineering, and medical communities in U.S. foreign policy and international development.

The JSF program is open to tenured, or similarly ranked, faculty from U.S. institutions of higher learning who are U.S. citizens. After successfully obtaining a security clearance, selected Fellows spend one year on assignment at the U.S. Department of State or the U.S. Agency for International Development (USAID) serving as advisers on issues of foreign policy and international development. Assignments may also involve travel to U.S. embassies and missions.

The deadline for applications is **October 16, 2020 at 5 PM ET**. To learn more about the program and to apply, visit www.nas.edu/jsf.

The JSF program is administered by the National Academies of Sciences, Engineering, and Medicine and supported by the U.S. Department of State and USAID.

The National Academies of
SCIENCES • ENGINEERING • MEDICINE

By Logan Brenner

An end and a beginning

The phone rang 30 minutes before my talk. It was a familiar number—my fertility doctor’s office—and I picked up, expecting a nurse to tell me that my pills and injections were ready for my next treatment. Instead, the nurse asked me to hold for the doctor, who told me I was pregnant. It should have been joyous news, but he quickly tempered my enthusiasm by telling me that, given the results of other tests, he wasn’t sure the pregnancy would go to term. I felt a complicated wave of emotions: relief that our fertility treatment worked, anxiety at the prospect of starting my new job pregnant, worry that I’d have a miscarriage—and stress about my imminent talk.

My husband and I had started fertility treatments during my postdoc years, unable to conceive on our own. The process was exhausting. I had to go into the doctor’s office a few times per week for blood tests and appointments—all before my hourlong commute to work. The costs stretched our finances. I began to convert every price tag I saw to a portion of a doctor’s visit.

During the same time period, my professional life had started to fall into place. I completed my postdoc and landed a faculty job in the same city where we were living. When the doctor’s office called that day, I was 1 month away from starting my new position and changing my email signature to “Assistant Professor.”

Somehow, I managed to give my talk. But as I delivered my practiced monologue, I was having a very different conversation in my mind. I tried to imagine taking maternity leave during my first year as a professor. I also fretted about the viability of the pregnancy.

Follow-up doctor’s visits didn’t yield positive news. The pregnancy would almost certainly end in a miscarriage. I contemplated termination. But because we’d had such a tough time getting pregnant, my husband and I decided to wait to see what happened.

Over the next month, my appointments for blood draws and invasive ultrasounds increased to nearly every other day. During one appointment, we heard a heartbeat. My stomach sank all the way down to the stirrups holding my feet. It was painful to hear that heartbeat—because it provided a glimpse into the excitement of pregnancy that I knew would likely be cut short. We left the office still waiting for the supposedly imminent miscarriage.

But I wasn’t waiting at home. I was waiting during my first faculty meeting. I was waiting in the lab while teaching interns how to prepare samples. I tried my best to



“It was a macabre preview into parenthood as an academic.”

silos my personal and professional lives. As time wore on, though, the barrier between “waiting for miscarriage Logan” and “shiny new professor Logan” became less and less distinct. I was worn out physically and emotionally.

It was a macabre preview into parenthood as an academic. I found myself constantly struggling to balance my career goals with my personal life, which featured many more doctor’s appointments than visits with friends. I missed deadlines and meetings. I had trouble focusing. And I cried in my office, worried that I was already failing as a professor—feelings intensified by hormones that would bear no fruit.

Then, roughly 7 weeks into my pregnancy, I went into the doctor’s

office—and this time there was no heartbeat. I’d never experienced such a visceral combination of sadness and relief. I swallowed two doses of the abortion pill and waited for my pregnancy to be officially over.

I dealt with my grief, in part, by speaking openly with trusted colleagues, including my postdoc adviser and others who’d known me for years. And I’d had time to prepare. Difficult as my experience was, I can’t imagine how someone feels when a miscarriage catches them totally off guard, or it happens for the second, third, or eighth time.

Ten months have passed, and my husband and I are still working with the fertility doctor. It isn’t easy to be open about my personal struggles. But I know many academics are in the same boat: starting their first faculty appointment while trying to start a family, and dealing with the conflicting emotions that even a better outcome can bring. I want them to know they’re not alone. ■

Logan Brenner is an assistant professor at Barnard College in New York City. Send your career story to SciCareerEditor@aaaas.org.