

A dark-horse company's race
for a COVID-19 vaccine p. 649


Enhancer function spans
sponges to mice pp. 657 & 681

Behavioral "nudges" reduce
arrest warrants pp. 658 & 682

Science

\$15
6 NOVEMBER 2020
SPECIAL ISSUE
sciencemag.org

 AAAS



SURVEYING BENNU

OSIRIS-REx investigates a
near-Earth asteroid p. 672

CONTENTS

6 NOVEMBER 2020 • VOLUME 370 • ISSUE 6517



654

NEWS

IN BRIEF

640 News at a glance

IN DEPTH

642 'A very, very bad look' for remdesivir

FDA and Europe anointed it as a key therapy just after a major study found it has little value

By J. Cohen and K. Kupferschmidt

644 Europe is locking down again—but its strategy is unclear

Without a plan to eliminate the virus, countries may face shutdowns over and over, scientists warn

By K. Kupferschmidt

645 Growth spurt for height genetics

By tallying thousands of genetic variants, researchers account for height's remaining "missing heritability"

By J. Kaiser

646 How the horse powered human prehistory

Wide-ranging warriors made Mongolian empire a melting pot, sweeping gene study shows

By A. Curry

647 New chip-based lasers promise practical terahertz imaging

Semiconductor lasers work with small coolers, enabling medical imaging and contraband detection

By R. F. Service

648 Black holes by the dozens challenge theorists

Trove of gravitational wave detections allows black holes to be studied en masse

By A. Cho

PODCAST

FEATURES

649 The long shot

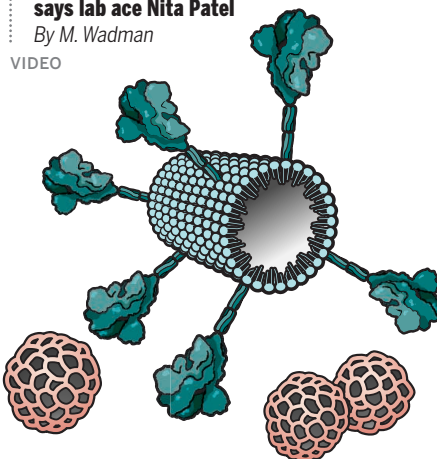
A little company chases its bigger competitors in the race for a coronavirus vaccine

By M. Wadman

652 'Nothing is impossible,' says lab ace Nita Patel

By M. Wadman

VIDEO



INSIGHTS

PERSPECTIVES

654 A surprising fossil vertebrate

An ancient amphibian converged on a chameleon-like way of feeding

By D. B. Wake

RESEARCH ARTICLE p. 687

655 The genetic law of the minimum

The genetic code evolved to reduce the impact of nutrient limitations

By M. F. Polz and O. X. Cordero

RESEARCH ARTICLE p. 683

657 Regulation in common: Sponge to zebrafish

Developmental enhancers from a sponge regulate gene expression in zebrafish neurons

By N. Harmston

RESEARCH ARTICLE p. 681

658 Nudging people to court

Behavioral interventions may improve court attendance, but how to reform the system remains unclear

By I. Kohler-Hausmann

RESEARCH ARTICLE p. 682

660 Food for thought

The nuclear metabolic-epigenetic axis bridges the environment and genes to modulate behavior

By G. Egervari et al.

662 Steering iceberg armadas

The Asian-Pacific tropics likely instigated millennial-scale climate changes

By J. M. Jaeger and A. E. Shevenell

REPORT p. 716

663 Public health during the pandemic in India

Data from two Indian states give detail about variation in transmission and disease outcomes

By J. John and G. Kang

RESEARCH ARTICLE p. 691

POLICY FORUM

665 Species protection will take more than rule reversal

Key improvements are needed for implementation of the Endangered Species Act

By Y.-W. Li et al.

BOOKS ET AL.

667 The ascent of Wikipedia

Scholars reflect on 20 years of crowdsourced knowledge

By A. Robinson

668 Recognizing the work of women

Despite several stumbles, a new volume about women in paleontology will likely prove valuable to future scholars

By J. Miller-Camp

LETTERS

669 Aquaculture jeopardizes migrating Oriental storks

By Y. Sun et al.

669 Indigenous rights to Patagonia

By P. A. Marquet et al.

670 Fisheries rely on threatened salt marshes

By R. Baker et al.

RESEARCH

IN BRIEF

677 From *Science* and other journals

REVIEW

680 Climate change

Past climates inform our future J. E. Tierney et al.

REVIEW SUMMARY; FOR FULL TEXT:

DX.DOI.ORG/10.1126/SCIENCE.AAY3701

RESEARCH ARTICLES

681 Enhancer genomics

Deep conservation of the enhancer regulatory code in animals

E. S. Wong et al.

RESEARCH ARTICLE SUMMARY; FOR FULL TEXT:

DX.DOI.ORG/10.1126/SCIENCE.AAX8137

PERSPECTIVE p. 657

682 Criminal justice

Behavioral nudges reduce failure to appear for court A. Fishbane et al.

RESEARCH ARTICLE SUMMARY; FOR FULL TEXT:

DX.DOI.ORG/10.1126/SCIENCE.ABB6591

PERSPECTIVE p. 658

683 Metabolic genomics

Resource conservation manifests in the genetic code L. Shenhav and D. Zeevi

PERSPECTIVE p. 655

687 Paleontology

Enigmatic amphibians in mid-Cretaceous amber were chameleon-like ballistic feeders J. D. Daza et al.

PERSPECTIVE p. 654

691 Coronavirus

Epidemiology and transmission dynamics of COVID-19 in two Indian states

R. Laxminarayan et al.

PERSPECTIVE p. 663

REPORTS

698 Organic devices

Large-area low-noise flexible organic photodiodes for detecting faint visible light

C. Fuentes-Hernandez et al.

701 Topological optics

Nonlinearity-induced photonic topological insulator L. J. Maczewsky et al.

705 Climate change

Global food system emissions could preclude achieving the 1.5° and 2°C climate change targets M. A. Clark et al.

708 Batteries

Rational design of layered oxide materials for sodium-ion batteries C. Zhao et al.

712 Ecological monitoring

Ecological insights from three decades of animal movement tracking across a changing Arctic S. C. Davidson et al.

PODCAST

716 Paleoclimate

Phasing of millennial-scale climate variability in the Pacific and Atlantic Oceans

M. H. Walczak et al.

PERSPECTIVE p. 662

721 Ecological genomics

Genomic architecture of a genetically assimilated seasonal color pattern

K. R. L. van der Burg et al.

725 Coronavirus

Free fatty acid binding pocket in the locked structure of SARS-CoV-2 spike protein

C. Toelzer et al.

DEPARTMENTS

639 Editorial

Gradually, then suddenly

By H. Holden Thorp

738 Working Life

A sabbatical reboot

By Thomas Gernon

Science Staff 636

Science Careers 734

SPECIAL SECTION

OSIRIS-REx at Bennu

INTRODUCTION

672 Sampling the early solar system

RESEARCH

674 Variations in color and reflectance on the surface of asteroid (101955) Bennu

D. N. DellaGiustina et al.

RESEARCH ARTICLE SUMMARY;

FOR FULL TEXT: DX.DOI.ORG/10.1126/

SCIENCE.ABC3660

675 Widespread carbon-bearing materials on near-Earth asteroid (101955) Bennu

A. A. Simon et al.

RESEARCH ARTICLE SUMMARY;

FOR FULL TEXT: DX.DOI.ORG/10.1126/

SCIENCE.ABC3522

676 Bright carbonate veins on asteroid (101955) Bennu: Implications for aqueous alteration history

H. H. Kaplan et al.

RESEARCH ARTICLE SUMMARY;

FOR FULL TEXT: DX.DOI.ORG/10.1126/

SCIENCE.ABC3557

ON THE COVER

The near-Earth carbonaceous asteroid (101955) Bennu, as observed by the OSIRIS-REx spacecraft. A lidar-derived shape model is overlain in grayscale with the albedo, and in color with a map of the abundance of carbon-bearing material (center, see page 675) and a false-color image indicating the near-ultraviolet and



near-infrared spectral slopes (right, see page 674). Graphic: NASA's Scientific Visualization Studio; Data: Daly et al. Sci. Adv. 6, eabd3649/ DellaGiustina et al./ Simon et al./ OSIRIS-REx Team

SEE ALSO SCIENCE ADVANCES RESEARCH

ARTICLES BY D. J. SCHEERES ET AL. 10.1126/

SCIADV.ABC3350, M. G. DALY ET AL. 10.1126/

SCIADV.ABD3649, AND B. ROZITIS ET AL.

10.1126/SCIADV.ABC3699

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 (51 issues): \$2148; Foreign postage extra: \$98. First class, airmail, student, and emeritus rates on request. Canadian rates with GST available upon request. GST #125488122. 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.sciencemag.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. Szurmi, Sacha Vignieri **SR. EDITORIAL FELLOW** Andrew M. Sugden (UK) **SR. EDITORS** Gemma Alderton (UK), Caroline Ash (UK), Brent Grocholski, Pamela J. Hines, Di Jiang, Marc S. Lavine (Canada), Yevgeniya Nusinovich, 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 Silks **LEAD CONTENT PRODUCTION EDITORS** Harry Jach, Lauren Kmec **CONTENT PRODUCTION EDITORS** Amelia Beyna, Jeffrey E. Cook, Chris Filiatreau, 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 Waldo (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** Lucy Hicks, Cathleen O'Grady **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 Piller, Michael Price, Tania Rabesandratana (Barcelona), Joshua Sokol, Emily Underwood, Gretchen Vogel (Berlin), Lizzie Wade (Mexico City) **CAREERS** Donisa Adams, Rachel Bernstein (Editor), Katie Langin (Associate Editor) **COPY EDITORS** Julia Cole (Senior Copy Editor), Cyra Master (Copy Chief) **ADMINISTRATIVE SUPPORT** Meagan Weiland

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 **DESIGN EDITOR** 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 **WEB DESIGNER** Jennie Pajewski

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 **PRODUCTION SPECIALIST** Kristin Wowk **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 Dhre

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 **PRODUCT DEVELOPMENT MANAGER** Scott Chernoff **PUBLISHING TECHNOLOGY MANAGER** Michael Di Natale **SR. PRODUCT ASSOCIATE** Robert Koepke **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 **US 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
SciencePodcast@aaas.org
ScienceVideo@aaas.org

INSTITUTIONAL SALES

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.sciencemag.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
Deji Akinwande, *UT Austin*
Judith Allen, *U. of Manchester*
Marcella Alsan, *Harvard U.*
Sebastian Amigorena,
Institut Curie
James Analytis, *UC Berkeley*
Trevor Archer, *NIEHS, NIH*
Paola Arlotta, *Harvard U.*
Johan Auwerx, *EPFL*
David Awschalom, *U. of Chicago*
Clare Baker, *U. of Cambridge*
Enad Ban, *ETH Zürich*
Franz Bauer,
Pontificia U. Católica de Chile
Ray H. Baughman,
UT Dallas
Carlo Beenakker, *Leiden U.*
Yasmine Belkaid, *NIAID, NIH*
Philip Benfey, *Duke U.*
Gabriele Bergers, *NIB*
Kiros T. Berhane, *Columbia U.*
Bradley Bernstein,
Mass. General Hospital
Joseph J. Berry, *NREL*
Alessandra Biffi,
Harvard Med.
Peer Bork, *EMBL*
Chris Bowler,
École Normale Supérieure
Ian Boyd, *U. of St. Andrews*
Emily Brodsky, *UC Santa Cruz*
Ron Brookmeyer, *UCLA (\$)*
Christian Büchel, *UKE Hamburg*
Dennis Burton, *Scripps Res.*
Carter Tribley Butts, *UC Irvine*
György Buzsáki,
NYU School of Med.
Blanche Capel, *Duke U.*
Annamarie Carlton, *UC Irvine*
Nick Chater, *U. of Warwick*
M. Keith Chen, *UCLA*
Zhiqian 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 Pitt.*
Roberta Croce, *VU Amsterdam*
Ismaila Dabo, *Penn State U.*
Jeff L. Dangi, *UNC*
Chiara Daraio, *Caltech*
Nicolas Daughas,
U. of Chicago
Christian Davenport,
U. of Michigan
Frans de Waal, *Emory U.*
Claude Desplan, *NYU*
Sandra Diaz,
U. Nacional de Córdoba
Ulrike Diebold, *TU Wien*
Hong Ding, *Inst. of Physics, CAS*
Dennis Discher, *UPenn.*
Jennifer A. Doudna, *UC Berkeley*
Raissa M. D'Souza, *UC Davis*
Bruce Dunn, *UCLA*
William Dunphy, *Caltech*
Christopher Dye, *U. of Oxford*
Scott Edwards, *Harvard U.*
Todd Ehlers, *U. of Tübingen*
Jennifer Eliseeff,
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*
Beate Escher,
UFZ & U. of Tübingen
Barry Everitt, *U. of Cambridge*
Vanessa Ezenwa, *U. of Georgia*
Michael Feuer,
GWU
Toren Finkel,
U. of Pitt. Med. Ctr.
Gwenn Flowers, *Simon Fraser U.*
Peter Fratzl,
Max Planck Inst. Potsdam
Elaine Fuchs, *Rockefeller U.*
Eileen Furlong, *EMBL*
Jay Gallagher, *U. of Wisconsin*
Daniel Geschwind, *UCLA*
Karl-Heinz Glassmeier,
TU Braunschweig
Ramon Gonzalez,
U. of South Florida
Sandra González-Bailón,
UPenn.
Elizabeth Grove, *U. of Chicago*
Nicolas Gruber, *ETH Zürich*
Hua Guo, *U. of New Mexico*
Kip Guy,
UK Coll. of Pharm.
Taejkip Ha, *Johns Hopkins U.*
Christian Haass,
Ludwig Maximilians U.
Sharon Hammes-Schiffer,
Yale U.
Wolf-Dietrich Hardt, *ETH Zürich*
Louise Harra, *U. Coll. 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*
Deirdre Hollingsworth,
U. of Oxford
Lora Hooper,
UT Southwestern Med. Ctr.
Fred Hughson, *Princeton U.*
Randal Hulet, *Rice U.*
Auke Huispeert, *EPFL*
Akiko Iwasaki, *Yale U.*
Stephen Jackson,
USGS & U. of Arizona
Erich Jarvis, *Rockefeller U.*
Kai Johnsson, *EPFL*
Peter Jonas, *IST Austria*
Matt Kaerberlein,
U. of Wash.
William Kaelin Jr.,
Dana-Farber Cancer Inst.
Daniel Kammen, *UC Berkeley*
V. Narry Kim, *Seoul Nat. U.*
Robert Kingston,
Harvard Med.
Nancy Knowlton,
Smithsonian Institution
Etienne Koehlin,
École Normale Supérieure
Alex L. Kolodkin,
Johns Hopkins U.
Julia Krupic, *U. of Cambridge*
Thomas Langer,
Max Planck Inst. Cologne
Mitchell A. Lazar, *UPenn.*
Wendell Lim, *UCSF*
Jianguo Liu, *Michigan State U.*
Luis Liz-Marzán, *CIC biomaGUNE*
Omar Lizardo, *UCLA*
Jonathan Losos,
Wash. U. in St. Louis
Ke Lu, *Inst. of Metal Res., CAS*
Christian Lüscher, *U. of Geneva*
Jean Lynch-Stieglitz,
Georgia Inst. of Tech.
Fabienne Mackay,
QIMR Berghofer
Anne Magurran, *U. of St. Andrews*
Asifa Majid, *U. of York*
Oscar Marin, *King's Coll. London*
Charles Marshall, *UC Berkeley*
Christopher Marx, *U. of Idaho*
David Masopust, *U. of Minnesota*
Geraldine Masson, *CNRS*
Jason Matheny, *Georgetown U.*
C. Robertson McClung,
Dartmouth
Rodrigo Medellín,
U. Nacional Autónoma de México
Graham Medley, *LSHTM*
Jane Memmott, *U. of Bristol*
C. Jessica Metcalf, *Princeton U.*
Baoxia Mi, *UC Berkeley*
Edward Miguel, *UC Berkeley*
Tom Misteli, *NCI, NIH*
Yasushi Miyashita, *U. of Tokyo*
Alison Motsinger-Reif,
NIEHS, NIH (\$)
Danielle Navarro,
U. of New South Wales
Daniel Nettle, *Newcastle U.*
Daniel Neumark, *UC Berkeley*
Beatriz Noheda, *U. of Groningen*
Helga Nowotny,
Vienna Sci., Res. & Tech. Fund
Rachel O'Reilly, *U. of Birmingham*
Harry Orr, *U. of Minnesota*
Pilar Ossorio, *U. of Wisconsin*
Andrew Oswald, *U. of Warwick*
Isabella Pagnold,
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. (\$)
Daniel Pauly, *U. of British Columbia*
Samuel Pfaff,
Salk Inst.
Julie Pfeiffer,
UT Southwestern Med. Ctr.
Philip Phillips, *UIUC*
Matthieu Piel, *Institut Curie*
Kathrin Plath, *UCLA*
Martin Plenio, *Ulm U.*
Katherine Pollard,
UCSF
Elvira Poloczanska,
Alfred-Wegener-Inst.
Julia Pongratz,
Ludwig Maximilians U.
Philippe Poulin, *CNRS*
Jonathan Pritchard, *Stanford U.*
Lei Stanley Qi, *Stanford U.*
Félix A. Rey, *Institut Pasteur*
Trevor Robbins, *U. of Cambridge*
Joeri Rogelj,
Imperial Coll. London
Amy Rosenzweig,
Northwestern U.
Mike Ryan, *UT Austin*
Shimon Sakaguchi, *Osaka U.*
Miquel Salmeron,
Lawrence Berkeley Nat. Lab
Nitin Samarth, *Penn State U.*
Jürgen Sandkühler,
Med. U. of Vienna
Olivia Eilamann Sapphire,
La Jolla Inst.
Alexander Schier, *Harvard U.*
Wolfram Schlenker, *Columbia U.*
Susannah Scott,
UC Santa Barbara
Rebecca Sear, *LSHTM*
Anuj Shah, *U. of Chicago*
Vladimir Shalae, *Purdue U.*
Jie Shan, *Cornell U.*
Beth Shapiro, *UC Santa Cruz*
Jay Shendure, *U. of Wash.*
Steve Sherwood,
U. of New South Wales
Brian Shoichet, *UCSF*
Robert Siliciano,
JHU School of Med.
Lucia Silvotti, *U. Coll. London*
Alison Smith, *John Innes Centre*
Richard Smith,
UNC (\$)
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 Sci.
V. S. Subrahmanian,
Dartmouth
Ira Tabas, *Columbia U.*
Sarah Teichmann,
Wellcome Sanger Inst.
Rocio Titunich, *Princeton U.*
Shubha Tole,
Tata Inst. of Fundamental Res.
Kimani Toussaint, *Brown U.*
Wim van der Putten,
Netherlands Inst. of Ecology
Reinhold Veugeler, *KU Leuven*
Bert Vogelstein, *Johns Hopkins U.*
Kathleen Vohs, *U. of Minnesota*
David Wallach,
Weizmann Inst.
Jane-Ling Wang, *UC Davis (\$)*
Jessica Ware,
Amer. Mus. of Natural Hist.
David Waxman, *Fudan U.*
Jonathan Weissman,
UCSF
Chris Wikle, *U. of Missouri (\$)*
Terrie Williams, *UC Santa Cruz*
Ian A. Wilson, *Scripps Res. (\$)*
Yu Xie, *Princeton U.*
Jan Zaenen, *Leiden U.*
Kenneth Zaret,
UPenn. School of Med.
Jonathan Zehr, *UC Santa Cruz*
Xing Zhu, *Inst. of Biophysics, CAS*
Xiaowei Zhang, *Harvard U.*
Maria Zuber, *MIT*

Gradually, then suddenly

Racism, climate denial, and coronavirus disease 2019 (COVID-19) are major crises standing in the way of a prosperous future for the United States, and resolution of all three could be enabled by science that is persistently ignored. In Ernest Hemingway's novel *The Sun Also Rises*, a character named Mike is asked how he went bankrupt. "Two ways," he answers. "Gradually, then suddenly." The resistance of U.S. policy to science has followed a similar path: It gradually built up over 40 years, beginning with the election of Ronald Reagan, but suddenly reached a tipping point in the chaos of 2020. Will the path to resolution also be gradual and then sudden, and if so, at what cost?

A saying incorrectly attributed to Winston Churchill holds that Americans always do the right thing but only after all other possibilities have been exhausted. Whatever the source, the idea lives on because it resonates and is no more apparent than in the failure of the United States to aggressively deal with 400 years of racial injustice. Slavery ended, but only after a civil war and decades of delay. The civil rights movement created important positive change, but only after civil rights leaders Dr. Martin Luther King Jr. and John Lewis boxed in President Lyndon Johnson so that he had little choice but to champion legislation or be associated by history with staunch segregationist George Wallace. Will people of color in the United States have to endure yet more violence from white supremacists before the next inflection toward racial justice?

As for confronting climate change, the prospects seem distant. Support for climate science has been steadily undermined by politicians catering to businesses dependent on fossil fuels and by religious conservatives suspicious of science because it argues for evolution. When California's Secretary for Natural Resources Wade Crowfoot challenged President Donald Trump on climate change, the president laughed and

said, "I don't think science knows, actually." Perhaps Trump knew he was saying something untrue but that many Americans agree with. Will wholesale environmental destruction have to occur before the United States does something about climate change?

When it comes to COVID-19, White House Chief of Staff Mark Meadows admitted, "We're not going to control the pandemic," making clear that Trump's only strategy is to wait for therapeutics and vaccines to soften the blow. Although prospects for both look promising, we are months, if not a year, away from reasonable supplies of either. In the same interview, Meadows said that we

would defeat the virus "because we're Americans." Such nationalistic exceptionalism is embarrassing. The virus doesn't "know" who is an American. Must hundreds of thousands more people die before the United States recognizes that humility in the face of challenge is the way to save lives?

Now that so many possibilities have been tried and exhausted, can science help push the country toward resolving these issues? Science must deal with the systemic racism that persists in our enterprise. There are scientifically sound measures that could promote greater racial justice in America, but the scientific community is in no position to advocate for racial justice if its own house is not in order, and that

requires difficult soul-searching about the underrepresentation of racial and ethnic groups as well as norms and practices of science that are not inclusive. Scientists must continue to speak out. Skepticism of the peril of COVID-19 has brought forth the response of science in ways never before seen. Scientists must hold on to that voice once the world gets past the pandemic. The old ideal of keeping politics out of science has not served the United States well. And scientists must continue to do the best science. Eventually, society will ask for help. Let's make sure science has the goods when they do.

—H. Holden Thorp



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

"Racism, climate denial, and COVID-19 are major crises standing in the way of a prosperous future for the United States..."

“We’re in for a whole lot of hurt. It’s not a good situation.”

Anthony Fauci, director of the U.S. National Institute of Allergy and Infectious Diseases, describing in *The Washington Post* the Trump administration’s disjointed, inadequate approach to controlling COVID-19 as U.S. infections rise.

IN BRIEF

Edited by **Jeffrey Brainard**



A man prepared to vote last week in a trailer outside of SoFi Stadium in Inglewood, California. Poll workers and early voters across the United States donned safety gear to complete balloting for the country’s elections before the final day of voting on 3 November. To reduce crowds, many election officials set up makeshift polling places.

Probe of virus origin begins online

COVID-19 | After an in-person planning visit in August, a long-awaited investigation into the origins of the pandemic coronavirus, organized jointly by the World Health Organization (WHO) and China, got off to a virtual start on 30 October, when Chinese and international teams exchanged information during an online meeting. They described progress tracing the virus to its presumed animal source, according to a statement released by China’s National Health Commission. But the session was private, and neither China nor WHO identified the scientists involved. *The New York Times* reported one American is on the WHO team, whose members China got to approve. The newspaper, citing WHO documents, also says the Chinese side will conduct the key studies in that country, with the WHO team reviewing them. Michael Ryan, WHO’s executive director for health emergencies, said at a 30 October press briefing “we fully expect” WHO investigators to visit China as part

of the effort, but he didn’t say when. U.S. President Donald Trump has criticized WHO for not confronting China about how it has handled the virus, but the agency has chosen to encourage cooperation, noting it has no authority to compel China’s assistance.

Contact tracing faces doubts

PUBLIC HEALTH | U.S. health officials tracking down people potentially exposed to the virus that causes COVID-19 likely confront several layers of discomfort and mistrust, according to survey results published last week by the Pew Research Center. The group asked 10,211 adult Americans in July how comfortable they were with various steps in the contact tracing process and how likely they were to participate. Only 48% said they would be comfortable or likely to engage in all three key steps: speaking to a public health official, sharing information about recently visited locations and contacts, and quarantining to prevent further

disease spread. The survey suggests a big hurdle is reaching people at all; 80% of participants said they don’t pick up phone calls from unknown numbers, and only 67% said they would check voicemail from such numbers.

Slovakia tries wide virus testing

COVID-19 | Slovakia last week tested more than half of its 5.5 million people for the pandemic coronavirus. The government hopes the ambitious program will avert the need for a second lockdown as cases surge in Europe. Of the 3.6 million residents tested last weekend, 38,359 tested positive and are required to self-isolate for 10 days or face a fine of €1650 (\$1918). Those who decline to be tested are also required to isolate. The antigen tests being used are faster but less accurate than polymerase chain reaction tests, prompting critics to warn that they may fail to diagnose up to 30% of positive cases, which risks letting infected people spread COVID-19. Another wave of mass testing is planned for 6–8 November.

NIH finalizes data-sharing rule

BIOMEDICINE | Starting in January 2023, all biomedical researchers newly funded by the U.S. National Institutes of Health (NIH) will be required to share data from their studies or explain why they can't. The new policy updates a 2003 requirement that covered only large projects with more than \$500,000 per year in funding. Now, all scientists submitting a grant application must include a plan for sharing data produced by their research, preferably in established repositories. NIH will allow certain exceptions, such as when ethical or legal reasons preclude disclosing anonymized information about people studied in clinical trials. Although a draft policy released 1 year ago merely urged "timely" sharing, the final policy sets a deadline: when a study is published or a grant ends, whichever comes first. NIH can cancel a grant or withhold future funding for failure to comply with the policy.

Millions sought for Arecibo repair

ASTRONOMY | Managers of the Arecibo radio telescope, the world's second largest, have asked the National Science Foundation (NSF) for at least \$10.5 million for repairs, United Press International reported this week. In August, a support cable slipped out of a socket and damaged part of the instrument's dish, which is nestled in hills in Puerto Rico. An investigation has not yet determined why the device failed, but as a precaution, managers may replace other cables at risk, which could bring the final repair bill past \$20 million, said Ray Lugo of the University of Central Florida, which manages the telescope for NSF, the owner.

Spherical fusion reactor turns on

ENERGY | The United Kingdom's revamped fusion reactor, known as the Mega Amp Spherical Tokamak (MAST) Upgrade, achieved its first plasma last week after a 7-year rebuild. The £55 million device may provide a stepping stone to an energy-producing facility. Doughnut-shaped tokamaks aim to generate energy by using powerful magnets to contain an ionized gas—or plasma—of hydrogen isotopes, heated by microwaves and particle beams to encourage fusion of the nuclei. Spherical tokamaks are shaped more like a cored apple than a doughnut, which researchers believe makes the roiling plasma more stable. MAST Upgrade first



CONSERVATION

United States removes gray wolf from endangered list

The Trump administration last week announced it is removing the protections of the U.S. Endangered Species Act for the estimated 6000 gray wolves (*Canis lupus*, above) living in the lower 48 states. The decision to give state and tribal agencies primary responsibility for wolf management reflects "the best scientific and commercial data available," Secretary of the Interior David Bernhardt said on 29 October. But many conservation groups disagree and say they will sue to overturn the delisting, which takes effect 4 January 2021. They argue that gray wolf populations in the lower 48 states, which some 45 years ago had dwindled to fewer than 1500 individuals, are not fully restored. The critics also contend that the administration has ignored research indicating that lifting federal safeguards could endanger some populations by allowing greater hunting. Courts have rejected previous efforts to entirely end federal protection for the wolves, and a new U.S. administration could seek to reverse the decision.

tested the concept on a large scale starting in 1999 and has now been upgraded with extra heating power.

U.S. withdraws from Paris accord

CLIMATE CHANGE | The United States this week was scheduled to formally withdraw from the Paris agreement to control greenhouse gas emissions, as President Donald Trump had announced in 2017. But "depending on the result of the U.S. election, this may be nothing more than a footnote in history," note Nathaniel Keohane and Kelley Kizzier of the Environmental Defense Fund, an advocacy group. If Democrat Joe Biden wins the presidency, he has said the United States, the world's second largest emitter, will immediately rejoin the pact. Other countries are expected to press a Biden administration to set an ambitious goal for reducing carbon. China and the European Union have recently promised emissions cuts deeper than their original commitments.

BY THE NUMBERS

300 million

Estimate of potentially habitable planets in the Milky Way, extrapolated from the 2800 exoplanets discovered by NASA's Kepler satellite and data from the European Space Agency about stars' brightness. The estimate applies to Earth-size planets that orbit Sun-like stars and are warm enough to have liquid water on their surfaces. (Upcoming issue of *The Astronomical Journal*)

16%

Share of U.S. patent applications in 2018 for inventions involving artificial intelligence, up from 9% in 2002. More than 60,000 applications were filed in 2018. (U.S. Patent and Trademark Office)



IN DEPTH

President Donald Trump and FDA Commissioner Stephen Hahn (right) met with Gilead CEO Daniel O'Day (left) after remdesivir received an emergency use authorization.

COVID-19

'A very, very bad look' for remdesivir

FDA and Europe anointed it as a key therapy just after a major study found it has little value

By **Jon Cohen** and **Kai Kupferschmidt**

October was a good month for Gilead Sciences, the giant U.S. manufacturer of antivirals. On 8 October, the company inked an agreement to supply the European Union with its drug remdesivir as a treatment for COVID-19—a deal potentially worth more than \$1 billion. Two weeks later, on 22 October, remdesivir became the first COVID-19 drug approved by the U.S. Food and Drug Administration (FDA). The decisions mean Gilead can cash in big in two major markets, both with soaring COVID-19 cases.

But they baffled scientists who have closely watched the clinical trials of remdesivir unfold over the past 6 months—and who have many questions about its worth. At best, one large, well-designed study found that remdesivir, which must be infused intravenously, modestly reduced the time to recover from COVID-19 in hospitalized patients with severe illness. A few smaller studies found no impact on the disease whatsoever, and none has found that the antiviral reduces patients' level of SARS-CoV-2, the causative virus. Then, on 15 October, the fourth and largest study delivered what some believed was a coup de grâce: The World Health Organization's (WHO's) giant Solidarity trial showed that remdesivir does not reduce mortality or the time COVID-19 patients take to recover.

Science has learned that both FDA's decision and the EU deal came about under un-

usual circumstances that gave the company important advantages. FDA never consulted a group of outside experts that it has at the ready to weigh in on complicated antiviral drug issues. FDA can tap that group, the Antimicrobial Drugs Advisory Committee (AMDAC), to review all available data on experimental treatments and then make recommendations about drug approvals—yet it has not convened once during the pandemic.

The EU, meanwhile, agreed to the remdesivir purchase price exactly 1 week before the disappointing Solidarity trial results came out. It was unaware of those results, but Gilead knew the trial was a bust; it had begun to review the data on 23 September.

"This is a very, very bad look for the FDA, and the dealings between Gilead and EU make it another layer of badness," says Eric Topol, a cardiologist at the Scripps

Research Translational Institute.

FDA has no obligation to convene the panel for its decisions, stresses AMDAC member David Hardy, an HIV/AIDS scientist of the University of California, Los Angeles. Yet Hardy is "amazed" the agency didn't consult it in this case. "This sets the standard for the first COVID-19 antiviral," he says. "That really is something that's very, very important."

FDA did not respond to *Science's* request to discuss why it opted against convening the committee, noting only that it is "at the discretion" of division directors. But FDA's inaction stands in sharp contrast to its handling of COVID-19 vaccines; it convened an advisory group to discuss potential dilem-

mas the very day of the remdesivir approval. Gilead, for its part, acknowledges it had seen an early draft of the Solidarity results before signing the EU deal. But Gilead has aggressively challenged the validity of the data, in part because the study was done in countries with widely varying health care standards.

That criticism has angered Solidarity investigators. Half the patients who received remdesivir were treated in Europe and Canada, WHO notes, and the others were not necessarily in countries with substandard health care. "It's appalling to see how Gilead tries to badmouth the Solidarity trial," says Marie-Paule Kieny, director of research at the French medical research agency INSERM and a former WHO officer.

HOPES THAT FADED

On 10 January, 2 days after China revealed SARS-CoV-2 causes COVID-19, researchers published an encouraging study in *Nature Communications* about remdesivir, originally developed to fight the Ebola and hepatitis C viruses. In both test tube and mouse studies, it had powerful inhibitory effects on a SARS-CoV-2 relative that causes Middle East respiratory syndrome. Two weeks later, doctors treated the first confirmed U.S. case with the drug and reported the 35-year-old man improved rapidly.

An interim analysis from a large-scale clinical trial by the National Institutes of Health (NIH), announced on 29 April, showed the drug reduced the median recovery time of severely ill, hospitalized COVID-19 patients from 15 days to 11 days. A second study, in China, appeared the same day and found no

Science's COVID-19 reporting is supported by the Pulitzer Center and the Heising-Simons Foundation.

statistically significant benefit, however. Two days later, FDA granted an emergency use authorization (EUA) that allowed the drug to be given for severe COVID-19, a move President Donald Trump praised in an Oval Office press event with Daniel O'Day, CEO of Gilead.

But the mixed messages about the drug continued: In August, a Gilead-sponsored study showed patients with moderate pneumonia treated for 5 days with remdesivir improved more quickly than those who received standard care, but oddly, those treated for 10 days did not. Nevertheless, shortly afterward, FDA expanded remdesivir's EUA to include all hospitalized COVID-19 patients.

To Topol, FDA's move resembled heavily criticized EUAs issued earlier for the malaria drug hydroxychloroquine—which the agency later rescinded—and antibody-rich “convalescent” plasma. In an open letter to FDA Commissioner Stephen Hahn, he wrote: “These repeated breaches demonstrate your willingness to ignore the lack of scientific evidence, and to be complicit with the Trump Administration's politicization of America's healthcare institutions.”

Many scientists expected WHO's Solidarity trial—which had 2750 patients on remdesivir, about three times as many as all other published trials put together—to resolve the drug's worth. Conducted in 30 countries, Solidarity compared remdesivir and three other repurposed drugs with each other and the standard of care. (Unlike the NIH and Chinese trials, it did not use a placebo.) None of the drugs lowered mortality among hospitalized patients, it found, and the investigators also noted that remdesivir did not affect “the duration of hospitalization” or whether COVID-19 patients required ventilators.

Solidarity described the results to FDA representatives on 10 October and in a preprint posted 5 days later. But 1 week after that, FDA approved the drug, having reviewed data only from the NIH study and two Gilead-sponsored trials. It had ignored the Solidarity data as well as the findings in China.

That infuriated the Solidarity team. “The mantra I've always heard as a joke about the FDA is that they say ‘In God we trust, everyone else has to provide data,’” Kieny says. “So look at all the data.”

Gilead Chief Medical Officer Merdad Parsey argued in an open letter posted the day of FDA's remdesivir approval that Solidarity “does not negate other study results—particularly from a trial designed with the strictest of scientific standards,” like NIH's study. The company further noted that the Solidarity results have not been peer reviewed and told *Science* WHO had yet to provide Gilead “the underlying data sets or statistical analysis plan” for the trial.

WHO says Gilead knew the analysis plan

before joining the trial and will receive the full data set once the study is complete, and that FDA traditionally reviews all available data, including unpublished findings.

Clifford Lane of the National Institute of Allergy and Infectious Diseases, who helped run the NIH study, says its main difference with Solidarity is “the degree of granularity” of the analyses of subgroups that may have benefited. “I think the Solidarity data are fine,” Lane says. “It's a very large study and it has a very robust endpoint.”

Richard Peto, a University of Oxford statistician and epidemiologist who helped design Solidarity and analyze the data, stresses that the WHO trial cannot prove remdesivir has no benefit for COVID-19. “Gilead and the FDA have sort of maneuvered us into a position where we're being asked to try and prove remdesivir does nothing rather than asking the usual way round, which is, ‘Can the manufacturers prove it does something?’”

It's still possible that remdesivir might help people at early stages of disease, says Martin Landray of Oxford, who is co-leading

complicated decisions given the imperfect nature of the data. ... The urgency of the clinical use gives all the more reasons to have an open discussion,” says Baden, whose group last met in October 2019.

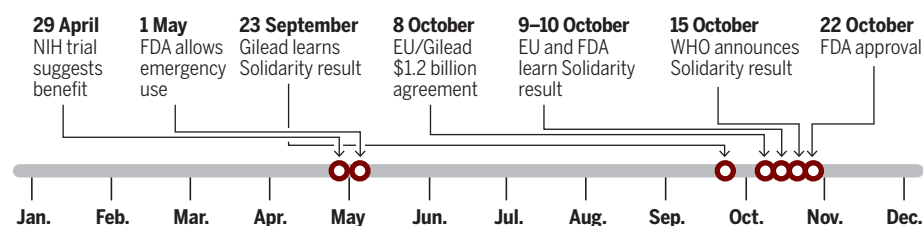
EUROPE IN THE DARK

EMA, Europe's FDA counterpart, in July gave “conditional approval” to remdesivir—similar to an EUA—but it has yet to give its full blessing. The European Union nevertheless has negotiated a “joint procurement agreement” with Gilead, signed on 8 October, that secures as many as 500,000 treatment courses for \$1.2 billion. A spokesperson for the European Commission confirms to *Science* it was not informed of the drug's failure in the Solidarity trial until the next day: “There was no discussion with WHO about the ongoing study prior to signing the contract.”

When *Science* asked Gilead why it didn't disclose the Solidarity data during the negotiations, the company said the draft manuscript it received from WHO in late September was “heavily redacted.” WHO says

Remdesivir's route to approval

The antiviral drug won U.S. approval for treating COVID-19 and a potentially lucrative deal with the European Union even as initial studies suggesting modest efficacy were followed by disappointing Solidarity trial data.



the world's largest study of various COVID-19 treatments. But it “definitely doesn't work in the sickest patients where the biggest gains would be.” Treating patients earlier comes at a price. “You won't save many lives, and you'll have to treat a lot of patients,” he says. “And it'll cost you a fortune.”

At the same time, the trials have not ruled out the possibility of harmful side effects. In late August, WHO noted a disproportionately high number of reports of liver and kidney problems in patients on remdesivir. And the European Medicines Agency (EMA) said last month that its safety committee had begun to assess reports of acute kidney injuries in some patients taking the drug.

To many scientists, such complexities underscore that FDA should have consulted AMDAC, its panel of outside experts, for a vigorous debate. It could have “elevated the discussion,” says AMDAC Chair Lindsey Baden, an infectious disease specialist at Brigham and Women's Hospital. “Hydroxychloroquine, convalescent plasma, remdesivir—these are

the only information blacked out was results relating to the other drugs used in the trial because of confidentiality agreements.

Although the agreement with Gilead locks EU members into a price of about \$2400 for a full course of remdesivir, it does not obligate any countries to purchase the drug, the Commission spokesperson says. But Yannis Natsis of the nonprofit European Public Health Alliance says that given the Solidarity results, the European Union “should at least renegotiate the volume of the doses and the price.”

To Kieny, this investment in a drug that may help just a few patients is an “enormous” waste. “You can always say, ‘OK, now, if I disaggregate the population and if I take only those who have a blue eye and a wooden leg, maybe this is very effective,’” she says.

On 28 October, Gilead told investors that remdesivir has brought in \$873 million so far this year. “We're proud to be at the front end of this with a very potent antiviral,” O'Day said. ■



A terrace in Quedlinburg remained empty after Germany's new lockdown took effect on 2 November.

COVID-19

Europe is locking down again—but its strategy is unclear

Without a plan to eliminate the virus, countries may face shutdowns over and over, scientists warn

By **Kai Kupferschmidt**, in Berlin

Shortly before 11 p.m. on 1 November, a waitress passed out paper cups to the customers crowded around the tables outside Luzia, a bar in Berlin's lively Kreuzberg district. "I'm sorry, but you all have to leave," she said. "God, in 2 minutes it's going to be lockdown," a woman at one table said, as guests poured the remainder of their cocktails into the cups. The fun was over: For the second time this year, Luzia had to close on the German government's orders.

All restaurants, bars, gyms, and theaters in Europe's largest economy will remain shut until at least the end of the month in a new bid to halt the spread of COVID-19. Hotels are no longer allowed to host tourists. Residents have been asked to meet people from only one other household. Florent, the manager at Luzia, took some hope from the fact that Germany was locking down while cases were still lower than in neighboring countries. "Hopefully we'll reopen in a month," he said.

With COVID-19 cases mounting and threatening to overwhelm health care capacity, much of Europe has taken similar measures to curb human contacts. Two months ago, as numbers began to creep up after a

blissful summer lull, countries still held out hope that more limited, targeted measures could prevent a second wave (*Science*, 4 September, p. 1151). Now, that wave is here, with the force of a tsunami. Europe has surpassed the United States in cases per capita; last week, it accounted for half of the more than 3 million cases reported to the World Health Organization (WHO). "Europe is at the epicenter of this pandemic once again," WHO's regional director for Europe, Hans Kluge, said on 29 October.

Most countries are reacting without a long-term plan, simply trying to avoid the worst. Officials differ about the best way to bring the numbers down again, and how low a level they should strive for. And no one knows what comes next. Short of vaccines to save the day, countries may face an exhausting series of lockdowns—a sawtooth pattern, "up and down and up and down," that could wreck the economy, says Albert Osterhaus, a virologist at the University of Veterinary Medicine, Hanover. "There is no strategy in Europe," he concludes.

The lockdown seemed an almost shockingly blunt tool when China first applied it in Hubei province on 23 January. But it also proved remarkably effective, and countries around the world took the same approach in

the spring, with varying degrees of intensity.

Europe has had a more science-driven pandemic response than the United States, but unlike many Asian countries, it was unable to avert a resurgence. Instead of using the summer to drive cases down to practically zero, Europe celebrated the holiday season. People seemed to lose their fear of the virus, says Michael Meyer-Hermann, a modeler at the Helmholtz Centre for Infection Research who was involved in drawing up Germany's lockdown plans. They increasingly flouted rules on physical distancing, mask wearing, and avoiding large gatherings.

"The infectious seeds in the community have always remained above a certain threshold, where if you relax the physical distancing, it's all going to come back," says University of Hong Kong epidemiologist Gabriel Leung. Numbers crept up and overwhelmed the other pillar of virus control: testing, isolating cases, and tracing and quarantining their contacts. The fact that life has largely moved indoors in recent weeks likely aided the resurgence, and cooler air may also favor the virus. "I think winter does make things much harder," says Adam Kucharski of the London School of Hygiene & Tropical Medicine. "Countries have probably been doing control on 'easy' setting over the summer."

Not everyone is convinced lockdowns are the answer. On 28 October, the day Chancellor Angela Merkel announced the new measures, Germany's National Association of Statutory Health Insurance Physicians presented a strategy paper arguing against a lockdown. "We cannot put the entire country, or even a continent, into an induced coma for weeks or months," said Andreas Gassen, head of the association. Another co-author, virologist Jonas Schmidt-Chanasit of the Bernhard Nocht Institute for Tropical Medicine, is convinced Germany's prelockdown restrictions would be enough to keep the virus from resurging—if they were strictly followed. Instead of closing tens of thousands of public places and spending billions to keep them afloat, Germany should spend money on communicating the need to better follow the rules, enforcing them strictly, and even creating some opportunities for people to safely enjoy riskier things such as parties, he says.

But most scientists say lockdowns are now unavoidable, although they need not be as draconian as in the spring, Kucharski says. Back then, "Countries were just doing absolutely everything at the same time," he says. Now, they can skip measures that severely restrict people's lives but don't contribute much to virus control. "There's no reason we need to lock people in their home," for

instance, as long as they keep their distance outside, says Devi Sridhar, chair of global public health at the University of Edinburgh.

The biggest difference from the spring may be that many countries are keeping schools open. Karl Lauterbach, a health policy expert and member of the Bundestag for the Social-Democratic Party who helped design Germany's "lockdown light," says it's clear schools do contribute to the spread of the virus. "Merkel asked: 'Am I sure we can do this without closing schools?' I said no, but probably yes," Lauterbach says. But the harm of school closures was deemed too great, he says.

Calculations suggest Germany needs to reduce contacts between people by roughly 75% from the current level, Lauterbach says. "That is incredibly hard if you want to keep schools and most businesses open." But bars and restaurants account for many contacts while providing only about 1% of Germany's gross domestic product, making them "kind of the perfect target for pandemic measures."

The first wave of lockdowns provided other lessons. For one, framing the debate as a choice between public health and the economy is wrong, Sridhar says: "In the longer term, having uncontrolled spread is much worse for the economy. That's what we've seen across the world." Also, locking down later means locking down longer. "If you wait until your infection level is quite high, probably your 2-week lockdown is going to become a 3-month lockdown," Sridhar says.

Europe's bigger question, however, is what comes next. Osterhaus says the strategy should be to get cases to zero using a hard and long lockdown, combined with strict border controls and quarantines to keep the virus out. That strategy, applied successfully by China, Australia, and New Zealand, may be politically unpalatable, however, and would require intense coordination between countries. "I can understand why that didn't happen," Leung says. "But the virus couldn't care less whether there are good or not so good reasons." Others are urging governments to follow South Korea's example: Drive cases to low numbers while improving the system of testing, tracing, and case isolation enough to keep the virus from resurging.

For now, Europe appears to be stuck in a third scenario: intermittently locking down to prevent the health system from buckling. "These circuit breakers are almost kicking the problem down the road," Kucharski says. "The hope is that vaccines and treatments and better options will be available in the new year." Lauterbach says he favored a longer lockdown in summer to go for elimination. But that would be too hard in winter, he says. "All we can do now is break the wave," he says. "But the water level will remain high." ■

BIOLOGY

Growth spurt for height genetics

By tallying thousands of genetic variants, researchers account for height's remaining "missing heritability"

By **Jocelyn Kaiser**

For height, DNA is largely destiny. Studies of identical and fraternal twins suggest up to 80% of variation in height is genetic. But the genes responsible have largely eluded researchers. Now, by amassing genome data for 4 million people—the largest such study ever—geneticists have accounted for a major share of this "missing heritability," at least for people of European ancestry. In this group, they've identified nearly 10,000 DNA markers that appear to fully explain the influence of common genetic variants over height.

"This is a genuine landmark," says Daniel MacArthur of the Garvan Institute of Medical Research in Australia.

If the missing genetic contributors to other traits and diseases can be identified, and extended to other ancestries, the results could "inform new biology and contribute to personalized medicine," suggests Loïc Yengo of the University of Queensland in St. Lucia, Australia, whose team presented the work online this week at the annual meeting of the American Society of Human Genetics. For example, geneticists could more accurately assess people's risk of diseases from genome scans. But the study leaves some scientists dissatisfied because it identifies only markers linked to genes that sway height, not the genes.

The mystery of missing heritability dates back to the late 2000s, when researchers began to use new tools to scan human genomes for common markers linked to diseases and traits. They expected the results of these genomewide association (GWA) studies to match evidence from the genetics of twins and families, such as the overwhelming influence in wealthy countries of genes on height, rather than environmental factors such as diet or childhood infections.

But it turned out that each identified marker, which may be in or just near a relevant gene, only contributes slightly to a trait or disease risk, and tallying them up didn't solve the problem. For height, the first 40 DNA markers linked to the trait explained just 5% of its variation.

A number of possible explanations emerged, including rare gene variants missed

by the GWA studies, gene-gene interactions, and that the twin studies were wrong. But Peter Visscher, leader of Yengo's team, argued it was partly a matter of finding many more common variants with very small effects. He estimated that such variants should account for 40% to 50% of the genetic component of height. Picking out the faint signals would require studying the DNA of a huge number of people, however.

By 2018, Visscher's team and other members of a global consortium called GIANT had pooled DNA data for 700,000 people and found 3300 common markers that explained 25% of the variation in height. Now, by looking across DNA from 201 GWA studies with 4.1 million participants, GIANT has brought the total to roughly 9900 common markers, accounting for 40% of the variation. Other markers located

nearby and likely inherited together account for another 10% of height variability.

That's still short of the 80% predicted by twin studies. But last year, Visscher's group drew on whole-genome sequencing data of a smaller number of people to demonstrate that

rare variants—those carried by fewer than one in 100 people—should explain another 30% of height's variation. (The result was released in a March 2019 preprint that the team is revising.)

Some geneticists say they aren't surprised that heritability gaps can be filled once enough people had their DNA scanned. "It was expected," says Aravinda Chakravarti of New York University. The problem remains that few of the height-linked DNA markers have been tied to specific genes that clearly alter the trait. "It's mostly all still 'missing' in a biological sense," says David Goldstein of Columbia University.

The new findings brighten prospects for tracking down those genes, Yengo says. Instead of being scattered randomly across the entire genome, which would make it challenging to ever get a handle on the specific genes and their roles, the 9900 or so markers are confined to only about 30% of the genome. And many cluster in areas known to contain genes involved in growth. ■

"This is a genuine landmark."

Daniel MacArthur,
Garvan Institute
of Medical Research



The horse was the heart of the mysterious Iron Age Xiongnu empire in Mongolia, as suggested by a decorative bronze belt plate showing two horses grappling.

ARCHAEOLOGY

How the horse powered human prehistory

Wide-ranging warriors made Mongolian empire a melting pot, sweeping gene study shows

By **Andrew Curry**

Until now, the only accounts of the Xiongnu came from their enemies. Chinese records from 2200 years ago describe how these fierce mounted archers from the wide-open steppes of today's Mongolia clashed with armies in what is now northwestern China. Their onslaughts spurred the Chinese to build what would become known as the Great Wall of China on their northern border, as protection against the mounted nomads. They also started to raise cavalry armies of their own.

The equestrian empire of the Xiongnu left no written records. But biology is now filling out their story, and those of other Central Asian cultures in antiquity. Two studies—a sweeping survey of ancient DNA from more than 200 individuals across 6000 years and an analysis of horse skeletons from just before the rise of the Xiongnu—trace population movements across Central Asia and the key role played by horsemanship. The results “show the horse was probably the driver of some of the ancestry shifts we see in the human population,” says Ludovic Orlando of Paul Sabatier University, who

was not involved in the paper. “The horse provided new range in patterns of human mobility and allowed people to travel long distance faster.”

Horses were probably domesticated by the Botai culture around 3500 B.C.E. near what is modern Kazakhstan (*Science*, 11 May 2018, p. 587). Horses may have been mainly used for meat and milk at first, and later began to pull wheeled chariots.

To learn more about human migration across Central Asia, a team led by Choongwon Jeong of Seoul National University and Harvard University's Christina Warinner sampled and sequenced DNA from human remains found in Mongolia. The results, which they report this week in *Cell*, span the period from 5000 B.C.E. all the way to the heyday of another horse-riding culture—that of Genghis Khan's Mongol Empire, around 1000 C.E.

Genetic studies of Western European populations have shown that around 3000 B.C.E., the Yamnaya—mobile herders of cattle, sheep, and goats—pushed west from the steppes of what is today Russia and Ukraine and triggered a dramatic genetic turnover in Europe. Skeletons from Bronze Age Mongolia had shown the Yamnaya also moved east and

introduced their dairy-oriented pastoralist lifestyle there. But they left no lasting genetic traces in Mongolia, the oldest samples in the new study show.

The ancient DNA does show that 1000 years later, another group from the steppes, called the Sintashta, left a lasting imprint. They also brought fateful cultural changes to Mongolia's grasslands, as earlier archaeological studies had shown. Starting in about 1200 B.C.E., equestrian innovations including selective breeding for size and endurance, plus bridle bits, riding pants, and even early saddles, appeared in the record, says archaeologist William Taylor of the University of Colorado, Boulder, a co-author on both papers.

Mongolians of the time were obviously riding horses, as vividly confirmed by the second paper, in the *Proceedings of the National Academy of Sciences*. The authors, Chinese and U.S. archaeologists, report that horse skeletons buried around 350 B.C.E. in the Tian Shan mountains, now part of China's Xinjiang province, show bone abnormalities from riding, including spinal damage from the weight of a rider and changes to the bones of the mouth from bits and bridles. “Put the lower back pathologies together

PHOTO: NATIONAL MUSEUM OF MONGOLIA

with evidence for a bridle, and it all suggests horses were being ridden,” says Sandra Olsen, an archaeologist at the University of Kansas, Lawrence, who was not part of either study.

Not long after, the Xiongnu emerged. They translated their skills on horseback into a sophisticated means of waging war and organizing an empire over vast distances. Starting in about 200 B.C.E., the Xiongnu marshalled nomadic tribes from across Eurasia into a formidable force, turning the steppes into a political center rivaling neighboring China. “The Xiongnu have been a source of constant worry and harm to China,” one contemporary Chinese historian wrote. “They move about in search of water and pasture and have no walled cities or fixed dwellings, nor do they engage in any kind of agriculture.”

Jeong’s study of DNA from 60 human skeletons from the Xiongnu’s 300-year-run shows how the region was transformed into a multiethnic empire. After more than 1000 years in which three distinct, stable human populations lived side by side on the Mongolian steppe, genetic diversity rose sharply around 200 B.C.E. Populations from western and eastern Mongolia mixed with each other and with people carrying genes from as far away as present-day Iran and Central Asia. Such wide-ranging mixing has “never been seen before at that scale,” Jeong says. “You can see the entire Eurasian genetic profile in the Xiongnu people.”

The results suggest mastery of the horse made possible stunning long-distance voyages on Central Asia’s sea of grass. Archaeological finds in the graves of Xiongnu elites, such as Roman glass, Persian textiles, and Greek silver, had suggested distant connections. But the genetic evidence suggests something more than trade. Eleven Xiongnu-period skeletons showed genetic signatures similar to those of the Sarmatians, nomad warriors who dominated the region north of the Black Sea, 2000 kilometers across the open steppe from Mongolia.

“There’s no written evidence of [Xiongnu] contact with Sarmatians, and it’s not well-attested archaeologically. It’s really surprising they’re mixing over these long distances,” says Tsagaan Turbat, an archaeologist at the Mongolian Academy of Sciences’s Institute of Archaeology. “This kind of information is really a game changer.”

In the future, researchers hope the genomes will help reveal how the mysterious nomad empire worked. The Xiongnu are “doing the things that empires do—forcing or enticing people to move,” says University of Michigan, Ann Arbor, archaeologist Bryan Miller. “Are people sent out to rule, or are local elites allowed to continue?” he asks. “Only genetics could answer that.” ■

TECHNOLOGY

New chip-based lasers promise practical terahertz imaging

Semiconductor lasers work with small coolers, enabling medical imaging and contraband detection

By Robert F. Service

Compact, chip-based lasers have conquered much of the electromagnetic spectrum, from ultraviolet to infrared, enabling technologies from digital communications and barcode readers to laser pointers and printers. But one key region of the spectrum remained untamed: the terahertz band, which lies between infrared light and microwaves. Engineers hankered for a ready source of terahertz radiation, which can penetrate opaque objects and probe chemical fingerprints inside. But compact terahertz lasers have only worked at ultralow temperatures, limiting them mostly to laboratory settings.

No longer. In this week’s issue of *Nature Photonics*, researchers report creating a grain-of-rice-size terahertz laser on a chip that operates at 250 K, or -23°C , a temperature reachable with a plug-in cooler the size of a cracker.

“This is a great achievement,” says Miriam Vitiello, a condensed matter physicist at the Nanoscience Institute of Italy’s National Research Council. “It has been a long-term goal in the community to push up the temperature of terahertz lasers,” she adds. “There is now a plethora of applications that can be done,” from medical imaging to explosives detection at airports.

Standard chip-based lasers generate photons when electrons fall into electron vacancies within a semiconductor alloy, whose makeup determines the color. Gallium nitride, for example, emits blue light, whereas gallium arsenide emits red. No semiconductor alloys emit photons in the terahertz range. (“Terahertz” refers to the light’s frequency: trillions of cycles per second.) In 1994, however, researchers at AT&T Bell Labs created a new kind of laser in which the semiconductor’s structure, not just its chemistry, determined the wavelength. Called a quantum cascade laser (QCL), it contained hundreds of layers of semiconductors of precise thicknesses. Electrons

injected into the structure cascade down hundreds of energy steps, shedding a photon at each one. Those photons were infrared in the first QCL, but in 2002 researchers in Italy and the United Kingdom created QCL lasers that emitted terahertz photons.

Those devices needed to be chilled to 50 K, but last year, researchers led by physicist Jérôme Faist at ETH Zurich unveiled a terahertz QCL made up of hundreds of alternating layers of gallium arsenide and aluminum gallium arsenide (AlGaAs) that works at 210 K. It still required bulky and expensive cryogenic coolers, however.

At higher temperature the electrons become unruly, leaping barriers between layers rather than cascading through the structure one step at a time. “Over-the-barrier electron leakage was the killer,” says Qing Hu, an electrical engineer at the Massachusetts Institute of Technology. So Hu and colleagues added more aluminum to the AlGaAs barriers to better confine the electrons. Hu’s team also had to prevent electrons from interacting in a way that caused them to leak through the AlGaAs barriers.

Now, the researchers have shown that by tailoring the

layered structure even more precisely—some layers were just seven atoms thick—they could make electrons behave at temperatures warm enough to be reached with standard compact thermoelectric coolers. What’s more, Hu says, the same strategy should enable the team to eventually make room temperature terahertz lasers.

Room temperature terahertz sources could be paired with terahertz detectors that also work at room temperature, which Vitiello and other researchers are now developing. That marriage could lead to technologies such as terahertz imagers able to distinguish skin cancer from normal tissue without a biopsy or watch airline passengers and cargo for hidden explosives, illegal drugs, and even pharmaceutical fakes. Faist says: “We have hoped for this for a very long time.” ■



A terahertz laser sits on top of a small cooler.

ASTROPHYSICS

Black holes by the dozens challenge theorists

Trove of gravitational wave detections allows black holes to be studied en masse

By **Adrian Cho**

As children once did before Christmas, astrophysicists have been eagerly awaiting the arrival of a catalog. This one lists not toys, but cataclysmic collisions of objects such as black holes—the ghostly ultraintense gravitational fields left behind when massive stars collapse. Last week, it arrived, courtesy of the world's network of gravitational wave detectors. “It’s very exciting!” says Selma de Mink, an astrophysicist at Harvard University. “There will definitely be a flurry of papers rushing to take the first stabs at the data.”

The observations come from three huge L-shaped optical instruments called interferometers that sense gravitational waves, the infinitesimal stretching of space triggered by the collisions. Two belong to the Laser Interferometer Gravitational-Wave Observatory (LIGO), instruments with arms 4 kilometers long in Louisiana and Washington that in 2015 first sensed the chirp of gravitational waves from two black holes spiraling into each other. Virgo, an interferometer near Pisa, Italy, with 3-kilometer-long arms, joined the hunt in 2017.

LIGO and Virgo had already spotted 11 events including a merger of two neutron stars, which shed light on how the universe forges heavy elements. Now, the team has cataloged 37 additional black hole mergers, another neutron star merger, and one possible merger of a black hole and neutron star. The catalog and statistical analyses, posted 28 October on the physics preprint server arXiv, “are superimportant,” says Carl Rodriguez, an astrophysicist at Carnegie Mellon University who was not involved in the work. “With an individual event, there’s only so much you can do in comparing to astrophysical models. But with a catalog you can not only begin to constrain the theory, you can start to understand the landscape.”

For example, by analyzing the 47 black hole mergers together researchers can probe how the black holes spin—a potential clue to how the pairs came together. If both black holes spin in the same sense that they orbit each other, they might have formed from a pair of stars born together in the same spinning cloud of gas. If one or both black holes spins in

a different direction—and especially if they spin in the opposite sense of the orbit—the black holes probably formed separately and paired up later.

It’s hard to tell how the spins of a single pair of black holes align from the burst of gravitational waves emitted during the merger, says Maya Fishbach, an astrophysicist and LIGO member from Northwestern University. But by analyzing the events en masse, scientists have teased out evidence that at least some of the mergers involve reversed spins, suggesting black hole pairs form in more than one way. “It seems like there might be multiple things going on,” Fishbach says.

Rodriguez has argued that the black holes could form separately in knots of old stars called globular clusters and then pair up. He notes that the overall rate of black hole mergers that LIGO and Virgo see seems to roughly match the rate his model predicts. “I shouldn’t toot my own horn—but I totally am going to,” he says. But he notes that because of the uncertainties, the data are also consis-

tent with such a mechanism producing only one-quarter of the mergers.

Analyses of all the events also show that when it comes to black holes, “the diversity is surprisingly large,” says Frank Ohme, a gravitational wave astronomer at the Max Planck Institute for Gravitational Physics. From details of the mergers’ chirplike signals, scientists can calculate the masses of the colliding black holes. They expected to find a “mass gap” between about 45 and 135 solar masses—the result of particle physics processes that should blow apart stars within a certain mass range before they can collapse into black holes.

However, LIGO and Virgo have spotted mergers involving black holes within the gap, including one weighing 85 solar masses. De Mink, who models the evolution of black hole pairs from binary star systems, says the mass gap is “such a clear prediction from the models that it’s hard to believe” it’s not there. Similarly, scientists expected another forbidden range below five solar masses, based not on theoretical modeling, but on observations of individual black holes peacefully orbiting normal stars. But one black hole falls below that limit.

Researchers can even probe how the number of black hole mergers may have changed over cosmic time, Fishbach says. The new data show the rate 8 billion years ago was no more than 10 times what it is now, she says, tightening the limit by a factor of 10,000.

Such data will challenge theorists as never before, says Ilya Mandel, a theoretical astrophysicist at Monash University. They can no longer simply whip up models to account for oddball events, but have to explain the entire set of observations. “When you have a whole population things become a lot harder,” Mandel says. “Your model not only has to reproduce the events that you do see, but also not predict events that you don’t see.”

The LIGO and Virgo teams owe their bounty to their detectors’ ability to spot ever fainter, more distant events. And they want to keep adding to it. “We’ve answered a lot of questions we didn’t even know we had,” Fishbach says, “but we raised even more. This is just the beginning of the science.” ■



Merging black holes might spin in different directions—a hint that they were born in separate star systems, then found each other.

IMAGE: LIGO/CALTECH/MIT/SONOMA STATE (AURORE SIMONNET)

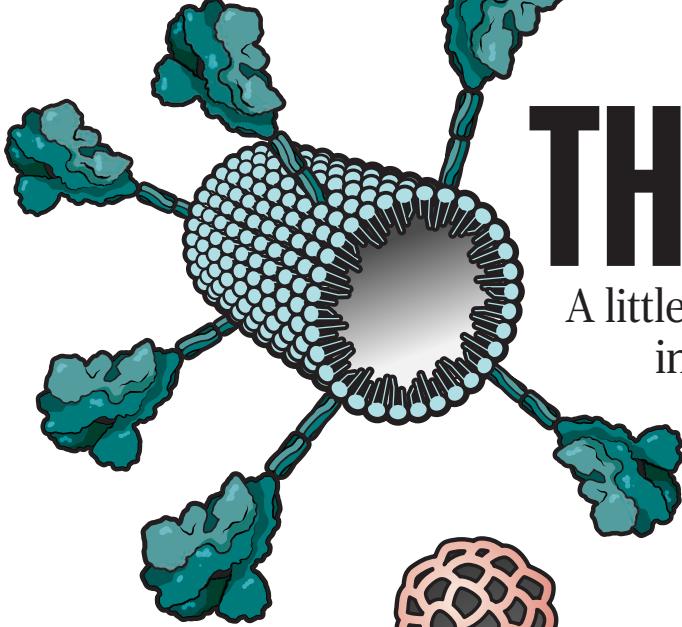
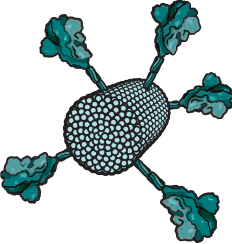




THE LONG SHOT

A little company chases its bigger competitors in the race for a coronavirus vaccine

By Meredith Wadman, in Gaithersburg, Maryland

Eighteen months ago, a small vaccine-maker here called Novavax faced an existential threat: delisting by the NASDAQ stock index. On the heels of a second failed vaccine trial in less than 3 years, the firm's shares had plunged to less than \$1 for 30 straight days, triggering a warning by NASDAQ. Frantic to conserve cash, the company sold its two Maryland manufacturing facilities, slicing its payroll by more than 100 employees. By January, it employed only 166 people.

"Good ideas. Bad management. ... The company will probably die soon," a former Novavax manager wrote on Glassdoor.com in October 2019.

What a difference a year—and a pandemic—make. Today, Novavax is slated to receive up to \$2 billion from the U.S. government and a nonprofit organization to develop and manufacture a coronavirus vaccine. The company's stock closed at \$80.71 per share on 30 October; it has hired more than 300 new employees, and this month it plans to launch a pivotal clinical trial of its coronavirus vaccine in the United States and Mexico. Made by moth cells harnessed to crank out the virus' spike protein—which the pathogen uses

to invade human cells—Novavax's vaccine outshone major competitors on key measures in monkey and early human tests.

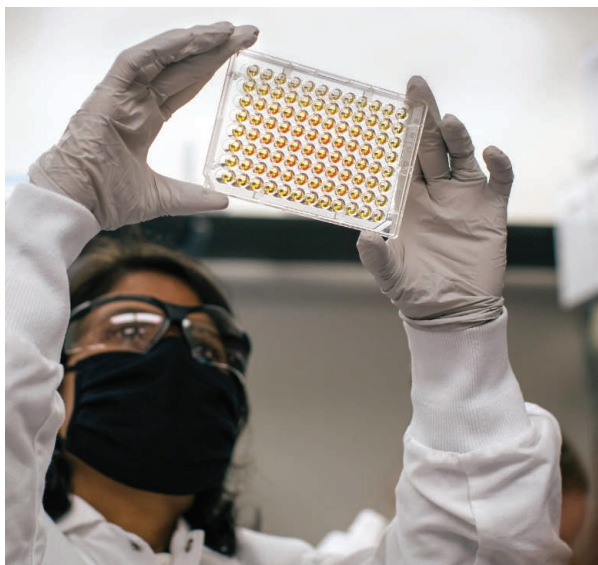
The company is one of just seven vaccine-makers to win funding so far from Operation Warp Speed, the giant multiagency U.S. government effort aiming to quickly produce at least 300 million doses of COVID-19 vaccines. But most Warp Speed-backed companies are giant pharmaceutical firms, and most have already launched late-stage clinical trials in the United States. Tiny Novavax is rushing to keep pace with its larger rivals because companies that win the first approvals from regulators will have big market advantages. Still, some observers say Novavax's technology gives it an edge.

"They are incredibly well positioned," says Andrew Ward, a structural biologist at Scripps Research. Ward, who receives no payments from the company but owns some stock, led a team that last month published

a paper in *Science* describing the structure of Novavax's tailor-made spike protein, the heart of its vaccine. He was impressed by its stability and conformation, as well as the vigorous antibody responses it has elicited in humans and animals. "They have the know-how," he says. "And they obviously, as we

Novavax's vaccine is made of tiny particles studded with the coronavirus spike protein plus honeycomblike molecules, derived from plants, that stimulate the immune system.

Science's COVID-19 reporting is supported by the Pulitzer Center and the Heising-Simons Foundation.



Novavax scientist Nita Patel (left) examines a plate measuring antibodies in vaccinated monkeys; other staffers gauge how the vaccine's protein binds to human cells.

confirmed, make a good product.”

But other people are skeptical. They note that Novavax has focused on making vaccines for more than 20 years but has never brought one to market, and that its senior executives have sold tens of millions of dollars of company stock since its share price began to soar this summer.

Most significantly, the company has an Achilles' heel. Novavax must rely mostly on contract manufacturers to meet its ambitious goal for 2021: producing enough vaccine to give 1 billion people two shots each. If manufacturing problems crop up—and the company last week said manufacturing delays had slowed launch of its late stage North American trial—competing vaccines may surge ahead. “That’s concerning,” says David Maris, a veteran drug industry analyst and managing director at Phalanx Investment Partners. Where small companies such as Novavax are concerned, he adds, “people do want to believe in fairy tales.”

ON 10 JANUARY, researchers in China published the genome sequence of the virus ravaging the city of Wuhan. Three days later, Gregory Glenn, president of R&D at Novavax, asked his staff to order from a supplier the gene for the virus' spike protein.

Glenn and other Novavax scientists had spent years developing “protein subunit” vaccines, so named because they employ a protein (or part of one) from the targeted virus, plus an immune-boosting compound called an adjuvant, to provoke an immune response. The company hadn't had a commercial success—its vaccine against a serious respiratory illness failed in clinical trials. But it had produced a promising flu vaccine aimed at older adults, which was nearing

the end of a pivotal trial. The company had also created protein subunit vaccines against two close cousins of the pandemic virus—the coronaviruses that cause severe acute respiratory syndrome and Middle East respiratory syndrome, using those viruses' spike proteins. Those vaccines hadn't made it to market, but Novavax had plenty of experience with the coronavirus family. Glenn believed it was his company's moment.

The gene for the spike protein was slow to arrive, however. Finally, at 6 a.m. on 3 February, a vice president from the supplier hand-delivered a red-capped vial bearing the gene to Novavax's beige brick building here. The virus still hadn't been officially named—the vial was labeled “Cov/Wuhan”—but Novavax was now out of the gate and in the race to tame it.

The company's scientists started to work “with frenetic pace,” Glenn says. Some of their competitors were already a lap ahead, working on their own vaccines. “There's no question [that we're] behind” several companies that also won Warp Speed funds, Glenn said on the morning of 24 September, the day Novavax launched its first phase III trial, of 15,000 volunteers in the United Kingdom.

Most of Novavax's key competitors—Moderna, Pfizer, Johnson & Johnson subsidiary Janssen, and AstraZeneca—had launched phase III trials by then. To make their vaccines, all four of those firms use new technologies based on genetic material that directs protein production, rather than delivering proteins directly. Those platforms rely on DNA loaded in disabled viruses or on messenger RNA to carry genetic instructions for building the spike protein. Cells within a vaccinated person then churn out the protein, alerting the immune system.

Developers of protein vaccines must develop their own version of the spike protein—one that closely mimics the naturally occurring spike and is stable enough to retain its immunological punch during manufacturing, packaging, and distribution. Most such vaccines include an additional compound called an adjuvant to help stimulate a strong, protective immune response. Those extra steps make protein vaccines slower to develop than those that deliver genetic instructions.

But protein-based vaccines also have a long track record of effectiveness, in contrast with the newer, largely unproven approaches. The successful hepatitis B vaccine licensed in 1986 and recommended for all U.S. babies in their first day of life is a protein subunit vaccine. So are a flu vaccine approved in 2013 and the human papillomavirus vaccines that have sent rates of cervical cancer plunging since the first ones were licensed in the 2000s.

Perhaps because the technology is tried and true, scores of other companies are also racing to develop protein subunit vaccines. Novavax is the only one to have launched a phase III trial. Of the other firms, the huge vaccinemaker Sanofi Pasteur is likely Novavax's biggest rival. It “is going to be formidable competition to the Novavax vaccine,” says Vijay Samant, a former head of vaccine manufacturing at Merck and now a consultant to vaccine companies. (Novavax is not a client.) Sanofi Pasteur has deep pockets, infrastructure, and experience, and markets vaccines against 19 infectious diseases.

But Novavax scientists say they're ready for the competition. “We've been getting ready for this our whole lives,” says Gale Smith, Novavax's chief scientist.

ONCE THE PANDEMIC coronavirus gene arrived in Gaithersburg, Maryland, on 3 February, the company spent weeks making more than 20 versions of the spike protein, aiming for a product as immunologically potent as possible. The winner was the most stable antibody-inducing protein, one that mimicked the energy-packed state of the spike just before it fuses with the host cell membrane.

In March, a team led by Nita Patel, a senior director in the vaccine development department (see sidebar, p. 652), confirmed in lab tests that the engineered protein bound tightly to its human cell-surface receptor. The results strongly suggested antibodies to Novavax's protein would interfere with the virus' own spike protein as it tried to fuse with cells.

Patel's boss, Smith, next enlisted Ward to verify the protein's structure and stability with electron microscopy. Other tests showed the Novavax spike is stable for many weeks at 2°C to 8°C—a key advantage over the Moderna and Pfizer vaccines, which need to be stored at -20°C and -70°C, respectively, and once thawed, last only days in the refrigerator.

Now, the challenge was to make the protein in the vast quantities that the world would need. Novavax had a system to do that, co-invented by Smith decades earlier and since used by the company to develop its other vaccine candidates: moth cells.

As a 34-year-old graduate student at Texas A&M University in 1983, Smith and colleagues had developed a system that could produce proteins in big quantities. The researchers started with an insect-infecting virus called a baculovirus, which had the virtue of a roomy genome that can accommodate large chunks of foreign DNA. The researchers inserted a gene for a human immune protein, interferon, into the virus and then used it to infect cells from the caterpillar form of a pest called the fall armyworm moth. The virus transferred the gene to the moth cells, which duly secreted human interferon.

Back then, editors of major journals had little interest in the discovery and repeatedly rejected Smith's paper, which found a home in an obscure new journal, *Molecular and Cellular Biology*. But today the system is widely used in biotechnology. Now, it is at work producing the Novavax vaccine at a plant owned by a contractor in Morrisville, North Carolina, and soon, it's expected, at other plants owned or contracted by Novavax in Europe, the United States, and Asia.

In North Carolina, countless baculoviruses loaded with the gene for the coronavirus spike protein are invading moth cells in

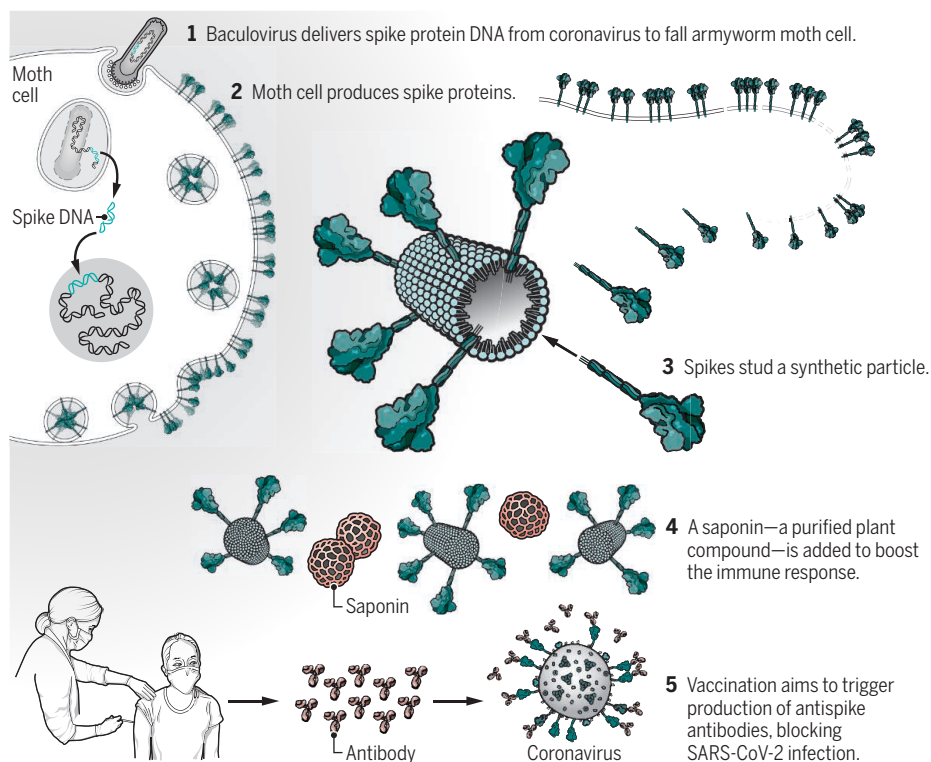
2000-liter bioreactors. The moth cells then express the coronavirus spikes on their cell membranes. Scientists harvest the proteins and mix them with a delivery vehicle: synthetic particles, on average 30 to 40 nanometers across. Each "nanoparticle" ends up studded with up to 14 spike proteins. The finished particles are only slightly smaller than the coronavirus itself, helping the immune system "see" them as a danger, Smith says. Then Novavax adds its adjuvant, based on saponin, a compound found in soap bark trees that stimulates the immune system (see graphic, below).

A vital piece of good news came on 24 March, boosting the company's search for cash. The phase III clinical trial of Novavax's flu vaccine in older adults, also made using the moth cell system, returned excellent results. For Novavax, which was now asking funders for major dollars for its pandemic coronavirus vaccine, the proof of its capabilities came at a fortuitous time.

Three days later, senior Novavax scientists met with decision-makers from a key funder—the nonprofit, international Coalition for Epidemic Preparedness Innovations (CEPI)—in a crucial Zoom call. Novavax sci-

The heart of a new vaccine

To make their vaccine, Novavax scientists first used a baculovirus to insert the gene for the SARS-CoV-2 spike protein into moth cells, which produced the spikes on their cell membranes. Scientists then harvested the spike proteins and mixed them with a synthetic soaplike particle in which the spikes embed. A compound derived from trees serves as an immune-boosting adjuvant.



To crank up production in the spring, Novavax had to cope with a harsh reality. "What we didn't have—what we'd lost by downsizing last year—is taking [the vaccine] right over to a manufacturing facility," Glenn says. "We could have been much earlier had we had that strategic asset." Instead, the company "had to beg" he says, turning to Gaithersburg-based contract manufacturer Emergent BioSolutions to produce vaccine for the first human trials. It was "not ideal," Glenn says. "We're sprinters ... [but] it's hard to get someone else to sprint."

tists were used to dealing with skeptics. "Most people don't think [our nanoparticles] work," Smith says. "We've been accused that [they] aren't even real, which is ridiculous." Then Ward's new high-resolution images—showing tiny particles studded with spike proteins, with honeycomblike saponins floating beside them—flashed on dozens of people's screens. It was "one of those precious moments," recalls Nick Jackson, CEPI's head of vaccine programs. "The excitement was palpable."

Six weeks later, CEPI awarded Novavax up to \$384 million for trials and manufacturing of its vaccine.

'Nothing is impossible,' says lab ace Nita Patel

Standing in front of reporters recently after showing Maryland Governor Larry Hogan (R) through the company's Gaithersburg, Maryland, labs, Novavax CEO Stanley Erck touted everything from the company's manufacturing plans to the long hours its scientists are logging. But he praised only one staffer by name: Nita Patel, a diminutive 56-year-old wearing jeans and a black mask. "Nita," Erck said to the governor, "has done all the work that you saw today."

Patel is a senior director in the vaccine development program at Novavax, a small firm among giant pharma companies racing to test a vaccine for the pandemic coronavirus (see main story, p. 649). Her all-female crew is an essential part of Novavax's lab team. Their sophisticated tests verified that the heart of the company's vaccine—its version of the virus' spike protein—performed as it should in cells and generated virus-neutralizing antibodies in animals. "Nita is absolutely invaluable," says her boss, chief scientist Gale Smith. "She's a genius."

Patel has come a long way from her beginnings in Sojitra, a farming village in India's Gujarat state. There, when she was 4 years old, her family fell into poverty after her father nearly died from tuberculosis (TB). He never worked again and told Patel she should become a doctor and find a cure.

Patel set about doing that, wearing the same ragged dress to school day after day. She had no shoes. She begged bus fare from a neighbor—at whose house she also devoured the newspaper because her family couldn't afford a subscription.

Her academic excellence propelled her through college on government scholarships. She later picked up two master's degrees, in India and the United States, in applied microbiology and biotechnology. Her memory is photographic: When driving, she has to be careful not to look at license plate numbers, or she'll memorize them.

Patel married a U.S. biochemist and then moved to Gaithersburg and started job hunting. One small company offered her less than others—but she would work on a TB project. In 1990, Patel became the 16th employee at MedImmune. One of her bosses there, Herren Wu, now a senior vice president at AstraZeneca, remembers her skill with a difficult assay that bedeviled others. "She was the one [whose data] I believed," he says. "She's a very good bench scientist."

But Patel also understands setbacks: A

MedImmune Lyme disease vaccine failed in its first clinical trial, and another therapy, against respiratory syncytial virus (RSV), was rejected by the Food and Drug Administration. In 2015, attracted by Novavax's RSV vaccine work, she jumped to the firm.

After Novavax got the gene for the SARS-CoV-2 spike protein in February, Patel's team tested more than 20 engineered variants of the protein, identifying the version most likely to elicit a protective immune response. Now, she's characterizing details of that protein, identifying the precise locations where neutralizing antibodies vigorously bind to it, and creating a test to ensure the spike is consistent from one manufacturing plant to another.

Since the pandemic arrived, she says, "my day just doesn't end. And it's the same with everyone else here." Yet Patel, who prays and meditates daily at a temple in her home, projects serenity and good cheer. "To me, nothing is impossible. So, having that mindset, nothing stresses me out, being honest."

Scientist Sonia Maciejewski, who works for Patel, agrees. "She has a very strong work ethic ... yet somehow doesn't put that sort of pressure or stress on us."

Patel's serenity gets a boost because she doesn't see the firm as competing with others. "We are [all] working towards, together, the world's problem," she says. —M.W.



Ward's work won over some scientific doubters. John Moore, an immunologist at Weill Cornell Medicine, had been skeptical of Novavax's moth cell system because in the 1990s it had conspicuously failed to produce an HIV spike protein with the right characteristics to make an AIDS vaccine. But in August, when Ward's work was posted as a preprint, "I looked at that paper and was impressed," Moore says. "It changed my perception of the quality of the protein. The concerns I had were eliminated by data, which is as it should be."

By late May, Novavax had launched its first human safety trial in 131 volunteers in Australia and used the CEPI funding to buy, for \$167 million in cash, a state-of-the-art vaccine manufacturing facility in the Czech Republic that the company said would deliver more than 1 billion doses in 2021. And in early July, Operation Warp Speed granted the company up to \$1.6 billion, with \$800 million available immediately, for a phase III clinical trial and for manufacturing 100 million doses of vaccine.

IN EARLY AUGUST, the big investors in the tiny company won an initial vindication when Novavax announced strong results from the Australian trial. After two injections, "the antibody responses in the Novavax paper were markedly stronger than any of the other vaccines that have been reported," and participants had experienced no severe adverse events, says Moore, who recently published a *Journal of Virology* review of the leading vaccine candidates. Moore says he intends to volunteer for a Novavax trial if eligible. "I'm going, 'Yes, I'll have that [vaccine].'"

The government of the United Kingdom soon signed up to buy 60 million doses of Novavax's vaccine, and the big drug-maker Takeda licensed it to manufacture at scale with funding from the Japanese government. Other scientists noted strong results in a dozen monkeys injected with various doses of Novavax's vaccine and then infected with live coronavirus. The virus failed entirely to multiply in the animals' noses and replicated in the lungs of just one monkey that received the lowest dose; that animal shut down the infection after 4 days.

"It's the only vaccine I've seen out of all the candidates that are further down the pipeline that actually had no viral replication in the nasal swabs of vaccinated animals," says Angela Rasmussen, a virologist at Columbia University. That's important, she says, because stopping viral replication in the nose can reduce the spread of infection among people who may be unaware they are sick. But she cautions that

monkeys are not people. “We can’t really conclude that this vaccine is going to be better in practice until we have some reliable safety and efficacy data in people.”

That’s why researchers will be eager to see results from Novavax’s phase III trials. In the one in progress in the United Kingdom, volunteers get either placebo or vaccine in two injections, 21 days apart. Researchers will evaluate whether vaccinated volunteers have fewer symptomatic coronavirus infections than placebo recipients. They’ll also keep monitoring safety, including any reactions to the adjuvant, because this will be the first time huge numbers of younger people with vigorous immune systems receive it.

Vaccine trials need volunteers who are exposed to the virus, so skyrocketing infections in the United Kingdom are likely working in Novavax’s favor. The company estimates there will be enough infections among participants to allow a first look at the data early in 2021—and says it expects the U.K. data will be enough to drive approval of its vaccine.

In the United States and Mexico, the company aims to enroll 30,000 volunteers—about 90% of them in the United States—in a trial expected to launch this month. Keeping volunteers in that trial may be challenging if other vaccine contenders win U.S. marketing approval in coming weeks (*Science*, 23 October, p. 389). But Glenn argues that this scenario is “not very realistic” given the time required for U.S. Food and Drug Administration (FDA) action and widespread distribution of a vaccine.

Observers also note that more than one vaccine will probably be needed to quell the pandemic. The first round of vaccine doses promised by Pfizer and Moderna combined would only cover 100 million people. Experts add that the first vaccine approved won’t necessarily be the best. “Especially when we are talking about vaccinating millions of people, you have just got to follow the science,” says Mayank Mamtani, a senior biotechnology analyst at B. Riley Securities. In the long run, “being first to market is just not, in my opinion, important.”

SOME OBSERVERS RAISED their eyebrows when Novavax won Warp Speed funding, given that the firm was on life support last year. Maris bemoans the “complete lack of

transparency as to how the funding decisions of Warp Speed were made.” (Journalists and investors have tried unsuccessfully to obtain Novavax’s Warp Speed contract.) “I would like to see the fundamental or deep analysis, if any, done prior to these funding decisions,” Maris says. “I’m not certain whether it exists.”

Novavax says it will release the contract next week as part of its quarterly financial reporting.

Other critics noted that since the pandemic began, Novavax top managers have made tens of millions of dollars by exercising options to buy their company’s stock at earlier, lower prices and sell it at this year’s

Novavax and two board members had pocketed nearly \$17 million in August alone.

Maris notes that executives who are bullish on their own companies tend to keep their stock. “So it’s interesting that they decided to sell.”

Novavax leaders continued their sales in September, with four senior managers selling shares worth a combined \$18.9 million. “Pretty much all my value I have is in Novavax,” says Glenn, who pocketed \$6.2 million (before taxes) in August and sold shares worth \$4.9 million in September. “A year ago it was worth nothing.” CEO Stanley Erck, who sold \$4.5 million in stock in September, did not respond to requests for comment.

Even if Novavax’s phase III trials succeed, vaccine watchers say the company’s biggest vulnerability will remain: the manufacturing process. “I think the vaccine is going to be efficacious,” Samant says. “But you put that 2 billion-dose task in front of a small company, it’s a huge uphill task.”

Ward concurs. “It’s like trying to cook a really good quiche. You can make one for your family, but if you try to make 50 of them for a wedding, that’s actually very difficult to do, to get them all the same and equivalent.”

“The [moth cell system] is not like the flu vaccine and eggs,” adds vaccine expert and FDA veteran Luciana Borio, an admirer of Novavax who is a vice president at In-Q-Tel, a technology investment firm. “It’s untested.”

Novavax counters that the company’s manufacturing partners are deeply experienced in making vaccines at scale. For instance, the Serum Institute of India, a veteran vaccinemaker that is the world’s largest, is helping Novavax expand production at

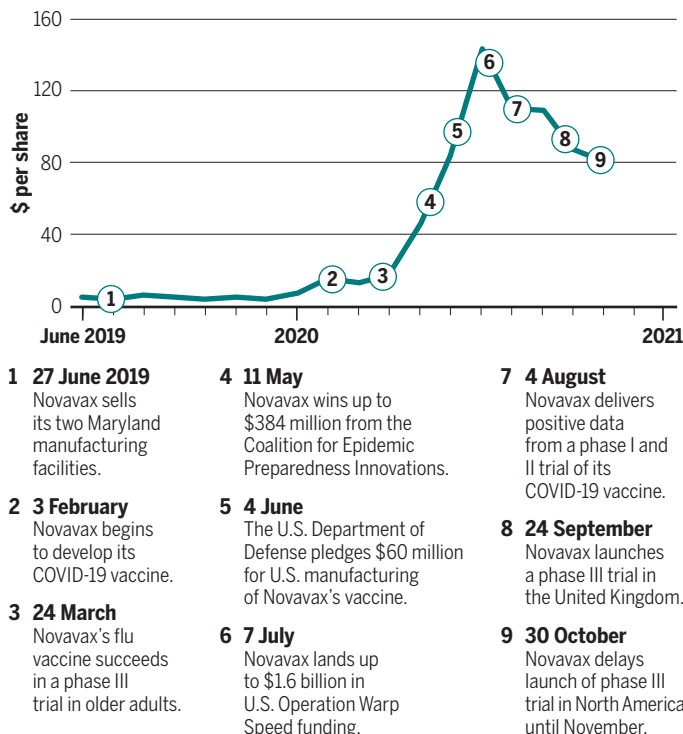
the 14,000-square-meter Czech facility, plus manufacturing an additional billion doses itself, in India, for low- and middle-income countries. That brings the total doses the company now says it can manufacture in 2021 to 2 billion, the promised number.

The fact that world-class manufacturers such as the Serum Institute “didn’t even blink before wanting to work with them and helping them scale up” speaks highly of the company, Mamtani says.

“It’s going to be a dark horse,” he predicts. “A dark horse helping us to get out of this darkness.” ■

Healthier prospects

In early 2019, Novavax’s share price fell so low that NASDAQ threatened to delist the company. Its candidate coronavirus vaccine has turned the small firm’s fortunes around, although its stock lost value in the fall.



high prices. Such options are a common way to pay biotech executives, but against the background of a pandemic, the sales led to an unwelcome spotlight on Capitol Hill in September. At a hearing, the chair of the House of Representatives Financial Services Subcommittee on Investor Protection, Entrepreneurship, and Capital Markets raised the examples of Novavax and other companies by name. And in a stinging memo that Democratic lawmakers issued under the heading “Corporate Integrity and Profiteering in the COVID-19 Pandemic,” they reported that three senior executives at

INSIGHTS

PERSPECTIVES



Ballistic tongue representation of *Yaksha perettii* is shown. The albanerpetontid is dwelling in the understory of the forest, and is illustrated an instant before being trapped by tree resin.

PALEONTOLOGY

A surprising fossil vertebrate

An ancient amphibian converged on a chameleon-like way of feeding

By David B. Wake

The invasion of land by vertebrates initiated an explosion of new kinds of organisms—amphibians—whose diversity ballooned until the extinction event that ended the Paleozoic Era 252 million years ago (Ma) nearly wiped them out. Several early amphibians became specialized in morphology and life history, even including forms that lost limbs.

Many had bizarre shapes, and they also varied greatly in size. Among them were the ancestors of the still-living salamanders, frogs, and caecilians, collectively known as lissamphibians. By the dawn of the Mesozoic Era, which followed the Paleozoic, only lissamphibians and one other group (trematosauroids) survived. The last trematosauroids disappeared in the late Mesozoic, 120 Ma. But paleontologists had overlooked one clade. On page 687 of this issue, Daza *et al.* (1) in-

troduce an unusual fossil of the obscure and apparently extinct albanerpetontids.

First detected in Late Cretaceous (100 to 66 Ma) formations and recorded as salamanders, albanerpetontids gradually became recognized by paleontologists as a fourth kind of lissamphibian, albanerpeton-

Museum of Vertebrate Zoology and Department of Integrative Biology, University of California, Berkeley, CA 94720, USA. Email: wakelab@berkeley.edu

ILLUSTRATION: STEPHANIE ABRAMOWICZ

tids (Order Allocaudata). With the discovery of several new taxa, albanerpetontids (named for the Albian, a mid-Cretaceous age and stage, 100 to 113 Ma) became better known. They were salamander-like, or maybe lizard-like, biologically, and persisted for a very long time, first appearing in the Triassic (252 to 201 Ma) and only disappearing “yesterday” (they were living in Italy with other still-surviving amphibians until the Early Pleistocene, about 2 Ma). With luck we might still have them somewhere (hopefully, one will turn up in what remains of biological wilderness on this planet—but time is running out). There is now a much better idea of the kind of organism we seek, thanks to the report by Daza *et al.* of their spectacular discovery of a new albanerpetontid, *Yaksha perettii*, encased in Myanmar amber (1).

The Myanmar amber, in this case about 99 million years old, preserves some spectacular fossils. In 2016, Daza’s team (2) reported amazingly well-preserved vertebrates, including one thought to be the earliest chameleon. The basis for that identification was a long skeletal element in the head and throat, an entoglossal, found only in chameleons and the key component in its intricate ballistic tongue. The very elongated entoglossal plays a special role in tongue firing in chameleons. It is enveloped by specialized accelerator muscles, which wrap around it. At the anterior end is the massive fleshy tongue pad. The entoglossal of chameleons slightly tapers toward the front, so when the accelerators squeeze down on it the tongue pad is swiftly advanced and slips forward off its end, literally flung into space (3). No other taxon was thought to have such an extreme entoglossal. Hence, the initial identification seemed to make sense. Logical—but spectacularly wrong. Daza *et al.* now show that not only are the bearers of the specialized tongue not chameleons, or even squamates; they instead are members of the mysterious albanerpetontids. What strange lissamphibians these albanerpetontids are: They have claws, scales, and armored skulls—and at least one had a tongue like a modern chameleon.

How the lissamphibians are related to each other and to Paleozoic forms is closely studied and contentious, but salamanders and frogs are likely sister taxa (Batrachia), caecilians are mysterious but might be derived from stereospondyls, and albanerpetontids seem likely to be basal derivatives.

Although many salamanders are almost prototypical tetrapods, with four limbs of roughly equivalent length, a generally long tail, and roughly proportional head and body, there are also some bizarre exceptions. Sirens lack hind limbs and are very elongate,

permanently aquatic forms with gills and a keratinous beak. Aggressively predaceous amphiumas are large, very elongate aquatic forms lacking gills. They have ludicrously small limbs and reduced digits—down to one in one species. Asian giant salamanders reach 180 cm in length and weigh over 50 kg. By contrast, frogs, the most numerous (88% of lissamphibians) and widespread living amphibians, are tailless, short-bodied, usually large-headed, and have limbs of often very unequal length. Caecilians are virtually blind, limbless, nearly tailless but very long-bodied burrowers, except for a bizarre clade of aquatic species in South American rivers and swamps.

Now, enter the albanerpetontids. Superficially salamander-like in their prototypical tetrapod habitus, albanerpetontids had features that seemed to suggest a burrowing way of life, with a strange, strengthened lower jaw with an interlocking articulation, a strengthened skull with solid bones covered with stout scales and elaborate bony sculpturing suggesting co-ossification of bone and skin, and with limbs bearing claws on their digits (four front, five hind). But what about the tongue? And how could these animals have been captured in amber sap? Were they arboreal (see the figure)? The closest analogs may be the strange tropical plethodontid salamanders of the genus *Oedipina*, a group of highly specialized burrowers, which—such as *O. allenii*—occasionally become arboreal. Their exquisitely specialized ballistic tongue is as if designed in reverse to that of chameleons—the homolog of the entoglossal (the basibranchial) is itself, together with the tongue pad, projected completely out of the mouth (4).

The elongated entoglossal is a vivid example of evolutionary convergence, which justifiably continues to receive intense scientific scrutiny (5, 6), because it is common and profoundly important. The expanded lissamphibians display different levels of convergence, including what might be termed complete convergence between *Yaksha* and chameleons, which entails both structural and functional components, and incomplete convergence between *Yaksha* and plethodontids like *Oedipina*, a convergence that is functional but only partially structural. ■

REFERENCES AND NOTES

1. J. D. Daza *et al.*, *Science* **370**, 687 (2020).
2. J. D. Daza *et al.*, *Sci. Adv.* **2**, e1501080 (2016).
3. D. E. Moulton, T. Lessinnes, S. O’Keeffe, L. Dorfmann, A. Goriely, *Proc. R. Soc. A Math. Phys. Eng. Sci.* **472**, 20160030 (2016).
4. S. M. Deban *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 10445 (2020).
5. D. B. Wake, M. H. Wake, C. D. Specht, *Science* **331**, 1032 (2011).
6. J. B. Losos, *Evolution* **65**, 1827 (2011).

10.1126/science.abe7826

EVOLUTIONARY BIOLOGY

The genetic law of the minimum

The genetic code evolved to reduce the impact of nutrient limitations

By Martin F. Polz¹ and Otto X. Cordero²

How organisms are optimized in the face of environmental challenges remains one of the key unanswered questions in biology. The optimization of enzymes for changing nutrient concentrations in an organism’s environment is well known (1). However, intricate genomics studies have revealed that optimization might affect the architecture of the entire genome (2). On page 683 of this issue, Shenhav and Zeevi (3) illuminate how selection driven by resource scarcity can affect the evolution of nucleotide and protein sequences in marine microbes.

In extremely nutrient-scarce regions of the ocean, microbial genomes are often small and streamlined, containing only the most essential genetic information (4). However, the essential selective factor often is not the absolute concentration of a single nutrient but rather its ratio with other required nutrients (5). For example, the sunlit ocean surface is typically limited in nitrogen but not organic carbon, because photosynthetic organisms require much of the former but produce copious amounts of the latter. When biomass from this surface layer dies and sinks, it is degraded by heterotrophic bacteria, which, because of the stoichiometry of elements in their food versus their cells, shift the balance toward carbon limitation (6, 7).

Recent work has shown that this transition from nitrogen to carbon limitation provides an explanation for an abrupt shift in the guanine-cytosine (GC) content in the genomes of marine microbes at the ocean surface versus those in the deep ocean (8, 9). Organisms living under consistent nitrogen limitation have a low GC content in their genomes, leading to a lower nitrogen demand for DNA synthesis. These orga-

¹Centre for Microbiology and Environmental Systems Science, University of Vienna, Vienna, Austria. ²Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. Email: martin.f.polz@univie.ac.at; ottox@mit.edu

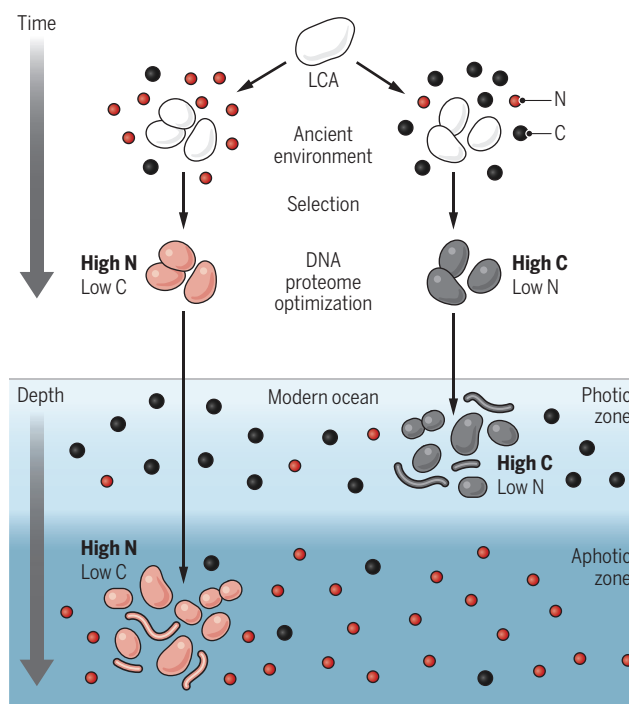
nisms also use codons that favor a proteome with a comparatively low nitrogen and high carbon content (10, 11).

Using a large set of metagenomic and single-cell genome data from across the global ocean, Shenhav and Zeevi show that natural selection purged nucleotide and amino acid changes that increase the demand for limiting resources. Selection was stronger at the protein versus DNA level, which is consistent with the high protein-to-DNA ratio in microbial cells. Highly expressed and secreted proteins showed the strongest signal for resource-driven selection because they place a high resource demand on the cell. Among the measured environmental variables, nitrate concentration—the prevailing form of nitrogen at the ocean surface (6)—was most strongly associated with the selection of protein sequences. Furthermore, mutations leading to lower nitrogen incorporation were inversely associated with mutations leading to lower carbon demand. This finding reflects the inverse relationship of carbon and nitrogen limitation as a function of water-column depth. The new insight that carbon limitation might also select for lower carbon incorporation led the authors to link these observations to the nature of the genetic code, which appeared to be simultaneously optimized for conservation of nitrogen and carbon. These findings strengthen the hypothesis that the microbial genome evolved to enable nutrient conservation in low-nitrogen and low-carbon environments.

Although variations of the standard genetic code (for example, in mammalian mitochondria or mycoplasma) show that the code itself can evolve, the variants are minor and derived from the standard code shared by all life-forms. Thus, the basic structure of the code, with its characteristic amino acid assignments, is believed to be ancestral to all extant cellular life-forms. Multiple (non-mutually exclusive) theories have been proposed to explain how the code originated (12). One of these—the adaptive theory—postulates that the genetic code evolved to maximize mutational robustness (that is, to lower the chance that random mutations lead to deleterious effects caused by translation errors) (13). The results of Shenhav and Zeevi add to this theory by showing that the code is also robust in the sense that it minimizes

Resource use is hardwired in the genome

Optimal resource use has been a selective evolutionary force since before the last common ancestor (LCA) diversified into modern life (top). The genetic code evolved to afford marine microbes the ability to adaptively minimize nitrogen (N) or carbon (C) content in their biomass, depending on the nutrient availability in the environment (bottom, ocean).



the impact of random mutations on carbon and nitrogen demand. Thus, the code seems to have evolved to minimize the potential damage of random mutations, both in terms of mistranslations and inefficient resource use. This new insight indicates that differential limitation by carbon and nitrogen constrained the evolution of the most primordial forms of life when the universal genetic code evolved.

The findings of Shenhav and Zeevi point to an incredible fine-tuning of genomes toward nutrient availability where selection acts on individual codons. However, whereas single mutations that affect the performance or regulation of proteins can have large fitness effects, the fitness gain of an individual-codon mutation in a protein that optimizes resource use must be vanishingly small. Using, as a reference, the composition of *Escherichia coli*, a cell with a single nitrogen-saving codon mutation in a highly expressed protein (present at 10^4 copies per cell), would experience a change in nitrogen demand on the order of only $\sim 0.001\%$, assuming $\sim 10^9$ nitrogen atoms per cell (14).

With these numbers in mind, it seems hard to imagine a fitness advantage large enough to cause the spread of an adaptive

mutation. It would therefore be illuminating to model the effects that such mutations might have on cell physiology and long-term fitness. For example, scientists could develop experimental models of competition experiments under nitrogen limitation to determine whether fitness effects are actually measurable (perhaps as a result of unanticipated pleiotropy). Such experiments could give researchers a better grasp of the magnitude of the relative fitness advantage experienced by the mutant organisms and the evolutionary dynamics that led to the patterns discovered by the authors.

The study by Shenhav and Zeevi reveals a fundamental selective force that has affected all forms of cellular life dating back to their last common ancestor and that seems to still be acting today on marine microbes. This work opens many new avenues of inquiry—for example, whether similar patterns of evolution are observed in other environments, such as soils or animal guts; whether mutational patterns of “amino acid choice” can be used to infer resource limitations from

metagenomic data; and whether purifying selection acts one amino acid at a time or whether other pleiotropic effects are required to explain the evolution of resource-driven codon selection in microbes. ■

REFERENCES AND NOTES

1. B. V. Adkar et al., *Nat. Ecol. Evol.* **1**, 149 (2017).
2. P. Baudouin-Cornu, Y. Surdin-Kerjan, P. Marlière, D. Thomas, *Science* **293**, 297 (2001).
3. L. Shenhav, D. Zeevi, *Science* **370**, 683 (2020).
4. S. J. Giovannoni, J. C. Thrash, B. Temperton, *ISME J.* **8**, 1553 (2014).
5. K. R. Arrigo, *Nature* **437**, 349 (2005).
6. C. E. Moore et al., *Nat. Geosci.* **6**, 701 (2013).
7. E. R. M. Druffel, S. Griffin, A. I. Coppola, B. D. Walker, *Geophys. Res. Lett.* **43**, 5279 (2016).
8. J. J. Grzymalski, A. M. Dussaq, *ISME J.* **6**, 71 (2012).
9. D. R. Mende et al., *Nat. Microbiol.* **2**, 1367 (2017).
10. J. G. Bragg, C. L. Hyder, *Proc. R. Soc. London Ser. B* **271**, S374 (2004).
11. P. Baudouin-Cornu, K. Schuerer, P. Marlière, D. Thomas, *J. Biol. Chem.* **279**, 5421 (2004).
12. E. V. Koonin, A. S. Novozhilov, *UBMB Life* **61**, 99 (2009).
13. C. R. Woese, *Proc. Natl. Acad. Sci. U.S.A.* **54**, 1546 (1965).
14. R. Phillips, R. Milo, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 21465 (2009).

ACKNOWLEDGMENTS

M.F.P. is funded by the Simons Foundation (LIFE ID 572792) and the U.S. National Science Foundation Dimensions Program (NSF 1831730). O.X.C. is funded by the Simons Collaboration on Principles of Microbial Ecosystems, award number 542395.

10.1126/science.abf2588

Regulation in common: Sponge to zebrafish

Developmental enhancers from a sponge regulate gene expression in zebrafish neurons

By Nathan Harmston^{1,2}

During the development of multicellular animals, distal gene regulatory sequences called enhancers are involved in determining when, where, and how much a gene is expressed (1). Enhancers contain transcription factor binding sites (TFBSs), and the combination of TFs bound determines the activity of an enhancer. There is a lack of understanding about how an enhancer sequence interprets regulatory state (input) to drive target gene expression (output), and how these sequences are constrained and/or modified over time. On page 681 of this issue, Wong *et al.* (2) identify a series of enhancers in a marine sponge (*Amphimedon queenslandica*) that respond to TFs expressed during zebrafish (*Danio rerio*) development. Given the large evolutionary distance between zebrafish and sponge (their common ancestor existed more than 700 million years ago), these enhancers lack detectable sequence homology with vertebrate genomes, yet their ability to function as enhancers in zebrafish demonstrates some kind of functional conservation without sequence conservation.

Throughout Metazoa, numerous pairs of genes have been identified that are conserved together, a phenomenon known as microsynteny, thought to arise primarily from cis-regulatory constraints (3, 4). By investigating gene pairs conserved across Metazoa, Wong *et al.* identified and tested a set of enhancers present within the intron of one member of a pair (the bystander gene), whose function is to regulate the expression of the other gene (the target gene) in the pair. Insertion of the sponge *Islet-Scaper* gene pair into zebrafish embryos revealed that *eISL*, an enhancer located in the intron of *Scaper*, drove the expression of *Islet* in neuronal cells, similar to the expression patterns of the zebrafish paralogs of *Islet*. Intriguingly, *eISL* could drive *Islet* expression in neurons, a cell type not present in sponges (see the figure). Therefore, this enhancer is responding to the set of TFs expressed in zebrafish neurons and driving expression of *Islet*. These findings are important for understanding both how enhancers

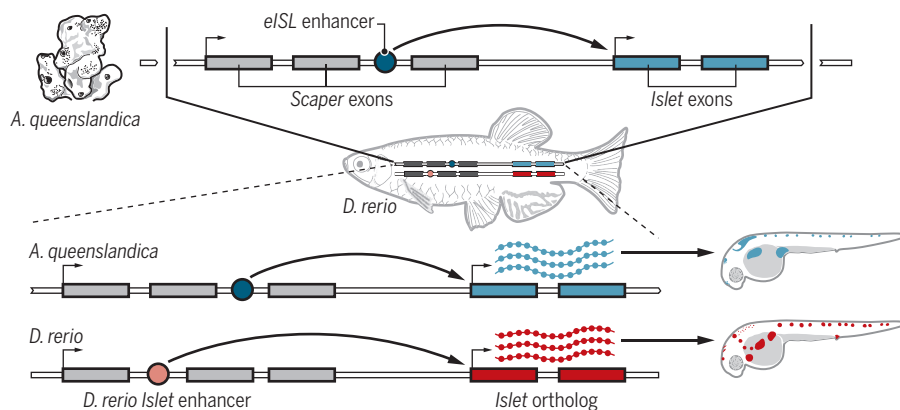
are structured and evolve, and how gene regulatory networks are reused and co-opted.

A large number of enhancers appear to be species-specific (5); however, there are a subset of enhancers that are conserved between evolutionary distant species. Comparing the genomes of different species within the same phyla has identified elements that show high noncoding conservation and which act as enhancers to regulate the expression of key developmental transcription factors (3). Mutations affecting these elements have been found to have substantial effects on pheno-

parable in terms of their constituent TFBSs (8). Together, these examples highlight that selection pressure on the output of some enhancers drives the conservation of enhancer sequence, whereas for other enhancers, sequence conservation does not appear to matter, and despite major differences in their organization, such enhancers are capable of generating the same output. Intriguingly, the set of enhancers identified by Wong *et al.*, though not conserved at the sequence level, can drive expression of their corresponding target gene in patterns similar to those of

Conserved enhancers

Injection of the entire *Scaper-Islet* locus of a marine sponge (*Amphimedon queenslandica*) into zebrafish (*Danio rerio*) embryos reveals that *eISL* (dark blue circle), a developmental enhancer located within an intron of *Scaper*, can drive *Islet* expression in zebrafish neurons. The *eISL*-driven gene expression patterns overlap with that of zebrafish orthologs of *Islet*.



types and in many cases are associated with developmental disorders (6). Therefore, small changes in enhancer sequence can have large effects on function and phenotype.

However, between closely related *Drosophila* (fruitfly) species, the content, ordering, and positioning of TFBSs (its cis-regulatory grammar) within the *eve* stripe 2 enhancer are different, yet all of the configurations lead to the same phenotype, indicating functional conservation despite a lack of sequence conservation (7). A brain enhancer that is functionally conserved between chordates and hemichordates has been identified. This enhancer can regulate the expression of sonic hedgehog (*Shh*) or hedgehog (*hh*) genes during brain development in mouse and acorn worm, respectively. Despite being distinct at the level of their overall DNA sequence, these brain enhancers are com-

parable in terms of their constituent TFBSs (8). Together, these examples highlight that selection pressure on the output of some enhancers drives the conservation of enhancer sequence, whereas for other enhancers, sequence conservation does not appear to matter, and despite major differences in their organization, such enhancers are capable of generating the same output. Intriguingly, the set of enhancers identified by Wong *et al.*, though not conserved at the sequence level, can drive expression of their corresponding target gene in patterns similar to those of

Multiple models have been proposed for how enhancers are structured in terms of the ordering and positioning of their TFBSs and the effect this has on the co-operative or additive binding of TFs to them. These range from the enhanceosome model, which requires the precise positioning of TFBSs (9), to the billboard model, where the order and positioning of TFBSs are not important (10), with the TF collective model allowing for cooperativity between TFs despite the lack of a well-defined grammar (11). Each of these models of enhancer organization implies distinct constraints on how enhancers evolve. Wong *et al.* used the set of TFBSs identified in *eISL* to search for similar sequences in other genomes and identified sequences in human, mouse, zebrafish, and fruitfly that had a simi-

¹Science Division, Yale-NUS College, 16 College Avenue West #01-220, 138527, Singapore. ²Programme in Cancer and Stem Cell Biology, Duke-NUS Medical School, 169857, Singapore. Email: nathan.harmston@yale-nus.edu.sg

lar TFBS content and were active in neuronal cell types. Although there was no apparent conservation between the entire sequences of *eISL* and these enhancers, short, conserved blocks of TFBSs were identified between them, indicative of a TF collective-like model for this enhancer. This indicates the existence of a common cis-regulatory grammar that is used across multiple lineages to regulate developmental gene expression in specific cell types, and furthermore suggests the presence of a regulatory state that has been reused or co-opted throughout evolution.

Together, these results indicate that there is a complex relationship between the structure and function of an enhancer and the gene regulatory networks that it is a part of, with each imposing constraints on the other (12, 13). The mismatch between their sequence and functional conservation raises questions about how to define when two enhancers are identical. In cases where there is a clear evolutionary history and similar function, it seems obvious that these enhancers are the same. But what if parts or the whole of an enhancer have turned over, resulting in a lack of observable sequence homology but still having the same function—is this still the same enhancer? Moreover, what techniques can be used to identify and characterize them? The *eISL* enhancer is not a true functional equivalent of the zebrafish enhancer, yet it is capable of driving overlapping patterns of gene expression. A critical question in understanding enhancer evolution is what causes some enhancers to require high sequence conservation to ensure conserved activity, whereas for others this does not matter (14)? Only through a combination of experimental and computational techniques is it possible to begin to systematically dissect the complex interplay of features affecting enhancer evolution to begin to fully understand the cis-regulatory grammar (15). ■

REFERENCES AND NOTES

1. I. Miguel-Escalada *et al.*, *Curr. Opin. Genet. Dev.* **33**, 71 (2015).
2. E. Wong *et al.*, *Science* **370**, eaax8137 (2020).
3. P. G. Engström *et al.*, *Genome Res.* **17**, 1898 (2007).
4. M. Irimia *et al.*, *Genome Res.* **22**, 2356 (2012).
5. D. Villar *et al.*, *Cell* **160**, 554 (2015).
6. L. A. Lettice *et al.*, *Hum. Mol. Genet.* **12**, 1725 (2003).
7. E. E. Hare *et al.*, *PLOS Genet.* **4**, e1000106 (2008).
8. Y. Yao *et al.*, *Nat. Genet.* **48**, 575 (2016).
9. D. Panne, *Curr. Opin. Struct. Biol.* **18**, 236 (2008).
10. M. M. Kulkarni, D. N. Arnosti, *Development* **130**, 6569 (2003).
11. G. Junion *et al.*, *Cell* **148**, 473 (2012).
12. J. L. Royo *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 14186 (2011).
13. P. Khoeiry *et al.*, *eLife* **6**, e28440 (2017).
14. N. Harmston *et al.*, *Philos. Trans. Royal Soc. B Biol. Sci.* **368**, 20130021 (2013).
15. T. Fuqua *et al.*, *Nature* **10.1038/s41586-020-2816-5** (2020).

ACKNOWLEDGMENTS

Thanks to E. Chua for her constructive comments.

10.1126/science.abe9317

SCIENCE AND LAW

Nudging people to court

Behavioral interventions may improve court attendance, but how to reform the system remains unclear

By Issa Kohler-Hausmann

New York City issues hundreds of thousands of summonses and misdemeanor arrests each year, highly concentrated in the city's poorer and minority communities (1, 2). On page 682 of this issue, Fishbane *et al.* (3) report that about 40% of the summonses (issued usually for infractions or violations that are classified below misdemeanors) issued in New York City in 2015 resulted in a failure to appear ("FTA") in court, which (usually) results in an arrest warrant and a host of collateral consequences (4, 5). To increase the likelihood that a person issued a summons shows up to the mandatory court date, Fishbane *et al.* experimented with a number of different behavioral "nudges," such as summons form redesign and texting different types of reminders to defendants. They found that all of these interventions decreased the FTA rate. It would be a positive step if other jurisdictions adopted these reforms. However, if the authors are correct that their study shows that FTAs are not exclusively caused by "choosing" not to appear in court, what role will that have in efforts to reform our massive and punitive criminal legal system?

Predicting behavior is distinct from understanding motives. Fishbane *et al.* suggest that the research design can adjudicate between two different motives for identical behavior: contempt for one's legal obligations versus lack of knowledge about one's legal obligations. This may be a mere framing device for the authors' true interest—perhaps simply establishing the existence of a treatment effect and estimating its magnitude. But behavioral economics infrequently so limits its explanatory ambitions. Rather, it aspires to illuminate underlying causal mechanisms by reference to general psychological principles. What makes such

topics interesting, yet difficult, is that observed behavioral propensities are usually compatible with multiple psychological explanations [e.g., (6)].

Although the evidence that the "nudge" treatments produce causal effects seems fairly straightforward, the authors' preferred explanation of the effect is less clear. What the authors present as opposing hypotheses do not seem to be mutually exclusive. The hypothesis the authors seem interested in refuting—that the high FTA rate for summonses is explained by defendants "deciding to skip court"—might capture many mental states. One such mental state might be "To hell with criminal court," where someone is aware of the specifics of their court obligations

and would be intent on noncompliance irrespective of more information about how to comply or the consequences of failure. Another might be "I can't afford to go to court today," where someone is aware of the specifics of their court date, but does not attend because they made a rational calculation that the benefits of nonappearance outweigh the costs (e.g., the opportunity costs of competing obligations to work or child care).

But one can "choose" not to attend court even while being unaware of the specifics of the obligation simply by "choosing" not to look at the information. Versions of this could include: "I realize this summons means I am supposed to go to court at some time, somewhere. But I feel hopeless and disillusioned from repeated injustice in the criminal legal system and I am tuned out," or "I am overwhelmed and busy trying to manage so many life obligations with so few resources, I never got around to this one." From my years studying and working in New York City's criminal courts, these seem especially plausible. But these do not seem conceptually exclusive of the authors' preferred hypothesis: "insufficient awareness" of legal obligations.

What I find more interesting, as a lawyer, sociologist, and reform advocate, is what these opposing theses say about how we think we must approach the politics of

"...they are not excused on the same terms many of us frequently invoke: We are busy ... overtaxed ... emotionally overwhelmed."

Yale Law School, New Haven, CT 06511, USA, and Department of Sociology, Yale University, New Haven, CT 06511, USA. Email: issa.kohler-hausmann@yale.edu



People wait to enter summons court on 346 Broadway in New York City, in 2011.

criminal legal reform. The hypotheses reflect our cultural assumptions: Either the people summonsed by the state—largely low-income Black and brown residents—are deserving of compassion or leniency because they are oblivious, or they are undeserving because they are obstinate. But they are not excused on the same terms many of us frequently invoke: We are busy, our time and resources are overtaxed, and we sometimes fail to discharge some of our obligations because we are just emotionally overwhelmed.

Unfortunately, the largely low-income and minority communities where summons enforcement is concentrated have had all too many encounters with the state's penal authority. Most recognize that the little piece of paper the police officer hands them represents another mandatory encounter with the state's coercive arm. The fact that it is confusing or mentally taxing to pin down the logistical details of how exactly to comply is distinct from the claim that one is not "aware" that the state has issued a sanction-backed demand to do so. To live in poverty and in highly policed neighborhoods in America is to be constantly subjected to such demands—from police, courts, welfare agencies, child services, landlords—and to hear a persistent message that one's failure to successfully perform is proof of one's unfitness for concern and respect in our polity.

At the height of "Broken Windows" policing in New York City around 2010, almost

three-quarters of summonses and half of misdemeanor arrests resulted in some kind of dismissal, most of them "earned" because defendants did community service, agreed to a temporary marking on their record, or just appeared in court until the prosecution was willing to dismiss charges (2, 7). Many New York City prosecutors hold "clean slate" events where anyone with an outstanding summons warrant can have it dismissed without adjudication or paying a fine. The entire logic of low-level enforcement is less about showing up to court for the purpose of adjudicating the underlying accusation than it is about showing up to court for the purpose of demonstrating that one has undertaken a cost to perform a responsibility.

I would respectfully disagree with the authors' assertion that our nation's punitive penal policies are "built on an assumption that people intentionally weigh the costs and benefits" of illegal conduct. I would say they are built on the assumption that the people targeted by such policies—largely people of color and impoverished people—are undeserving of full and equal membership in our political community absent a positive showing of worthiness.

Consider the evidence presented by Fishbane *et al.* in which MTurk participants responded to information about hypothetical "failures to take a required action," such as paying bills, turning in educational paperwork, and making a required criminal court appearance (i.e., FTA). Compared to

most other domains, respondents rated court FTAs as less likely attributable to "forgetting" and more likely to be "intentional." And respondents were least supportive of "nudges" in the criminal justice domain. These findings might just be ascriptions of blameworthiness. When people are asked how likely it is that a criminal defendant "intentionally and deliberately decided to skip their appointment," they might be reporting how likely they think it is that this person is as culpable as "intentionally deciding" to skip, whether or not the person actually knew when and where the appointment was or remembered its occasion (8). Similarly, respondents' expression of preferences for punitive measures over "nudges" in response to court FTAs does not necessarily mean they believe one is more efficacious than the other in the criminal justice domain. It might well reflect their political moral judgments that people who miss court, irrespective of the reasons, deserve punishments and not nudges. This is especially likely given evidence that whenever you ask someone about their attitudes toward crime, they are often also expressing their attitudes toward race (9, 10).

Although it would be positive if many jurisdictions would adopt the practices described by Fishbane *et al.*, changing the approach to penal and welfare policy in our country will require interventions that are much more radical than cost-neutral behavioral nudges that everyone can agree on. ■

REFERENCES AND NOTES

1. New York City does not collect racial identification data on the summons form. Data on misdemeanor arrests show that about 80% of fingerprintable misdemeanor arrests in recent years are of Black or Hispanic individuals. See (2).
2. I. Kohler-Hausmann, *Misdemeanorland: Criminal Courts and Social Control in an Age of Broken Windows Policing* (Princeton Univ. Press, 2018).
3. A. Fishbane *et al.*, *Science* **370**, eabb6591 (2020).
4. The FTA rate for misdemeanor arrests (including "desk appearance ticket" arrests) is substantially lower than 40%, with about half of those returning voluntarily within 30 days of the missed court date. See data from (5).
5. New York City Criminal Justice Agency, CJA Annual Report of 2017 (pp. 19, 36–38); <https://www.nycja.org/publications/cja-annual-report-2017>.
6. D. Gal, D. D. Rucker, *J. Consum. Psychol.* **28**, 497 (2018).
7. P. Chauhan *et al.*, The Summons Report: Trends in the Issuance and Disposition of Summonses in New York City 2003–2013; www.jjay.cuny.edu/sites/default/files/news/Summons_Report_DRAFT_4_24_2015_v8.pdf.
8. This elision is what is behind much of criminal law doctrine on "willful ignorance," where courts and juries find defendants guilty because they believe choosing not to know under certain circumstances is just as culpable as knowing.
9. R. C. Hetey, J. L. Eberhardt, *Psychol. Sci.* **25**, 1949 (2014).
10. This is analogous to a situation where people responding to the question "Do workers who are late for work because of childcare problems deserve to be fired?" might also be expressing beliefs about women in the workforce.

Published online 8 October 2020

10.1126/science.abc2495



EPIGENETICS

Food for thought

The nuclear metabolic-epigenetic axis bridges the environment and genes to modulate behavior

By Gabor Egervari, Karl M. Glastad, Shelley L. Berger

To drive behaviors essential for survival, numerous pathways have evolved that translate environmental stimuli into gene expression. These include interactions between metabolic and epigenetic regulation, which encompass both predictable indirect connections and unanticipated direct contacts (1). Indirect pathways involve metabolites generated from diet and other sources that provide substrates and cofactors for epigenetic enzymes to ensure seamless adaption to nutrient availability. In addition, epigenetic mechanisms regulate the expression of metabolic enzymes, which can in turn produce or deplete various metabolites. The direct metabolic-epigenetic interface, however, represents a paradigm shift in our mechanistic understanding of environmental impacts on gene expression and behavior. Direct communication is mediated by metabolic enzymes—classically thought to reside in mitochondria and cytoplasm—localizing within the nucleus and even binding to chromatin. This surprising relocation enables the generation of metabolite pools that can fuel epigenetic enzymes directly.

Classic examples of indirect metabolic-

epigenetic interactions are the nuclear receptor superfamily, which bind circulating small molecules, such as steroid and thyroid hormones. The binding of their ligands unleashes nuclear receptors to bind to DNA as transcription factors. They subsequently recruit coactivators and corepressors involved in chromatin modulation, resulting in the activation of new transcriptional programs. These mechanisms translate changes in circulating hormones into specific physiological and behavioral outcomes (2). An intriguing alternative model proposes that circulating hormones regulate the activity of certain epigenetic enzymes and thereby alter gene expression and behavior. Specifically, estradiol can increase brain masculinization in mice and rats by reducing DNA methyltransferase activity, which prevents or reverses DNA methylation. This results in de-repression of masculinizing genes in sexually dimorphic brain regions during neonatal development (3). Of note, pharmacologic or genetic inhibition of DNA methyltransferases recapitulates the effect of gonadal steroids, resulting in masculinized neuronal markers and male sexual behavior in female rats (3).

Conversely, epigenetic pathways can determine amounts of circulating hormones, leading to persistent developmental and behavioral changes. Relevant model organisms are ants and other social insects—bees, some wasps, and termites—wherein closely related individuals exhibit dramatic differences in

The development of carpenter ants (*Camponotus floridanus*) into different worker castes is regulated by the metabolic-epigenetic axis.

morphology and behavior. For example, in addition to the distinction of the female reproductive queen from sterile workers, some ant species have evolved multiple female worker classes. These so-called castes specialize in either foraging or defense behaviors that are vital for the survival of such complex societies. Recent findings indicate that epigenetic-to-metabolic signaling pathways play a critical role in caste determination in carpenter ants (*Camponotus floridanus*). Juvenile hormone (JH) is pivotal to programming foraging, and repressive epigenetic pathways expressed during brain development establish low expression of enzymes that degrade JH to preserve JH in the adult brain (4). Transient experimental manipulation of the brain epigenome in early adulthood of the defense caste promotes long-lasting behavioral reprogramming through stable epigenetic repression of the JH-degrading enzymes, switching behavior to foraging (4). Thus, the epigenome “encodes” metabolic information either naturally expressed during development or transiently encountered during vulnerable windows in early life, leading to long-term ossification of adaptive behavioral states. Many social insects can drastically alter their behavioral repertoire over their life span to satisfy colony demands. The precise role of metabolic-epigenetic reprogramming in the regulation of these phenomena later in life is currently unknown.

Beyond hormonal regulation, intermediary metabolism also influences epigenetic mechanisms. Metabolites serve as substrates and cofactors for epigenetic enzymes, and their availability and concentration can alter gene expression (1). Furthermore, provocative recent findings identified previously unknown histone modifications that can be driven by metabolites. These include, for example, the direct incorporation of lactate into chromatin. Lactate covalently binds to several lysine residues on core histone proteins, in a process potentially regulated by histone acetyltransferases, such as p300 (5). Histone lactylation is induced by hypoxia and by bacterial infection in human and mouse cells, helping to reestablish homeostatic gene expression (5). Further examples include the deposition of bioamine neurotransmitters serotonin (6) and dopamine (7) as covalent modifications of histones. Histone serotonylation potentiates binding of transcription factor IID to methylated histones (6). Serotonylation is enriched in the brain and gut (6), which are the primary sites of serotonin production, raising the intriguing possibility that incorporation into chromatin is dependent on metabolic

Epigenetics Institute, Perelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA. Email: bergers@penmedicine.upenn.edu

availability in those tissues. Similarly, histone dopaminylation occurs in dopamine-producing neurons of the ventral tegmental area, a midbrain region involved in reward and motivated behaviors. Dopaminylation accumulates during cocaine withdrawal, leading to transcriptional changes in this brain region. Inhibiting histone dopaminylation rescues cocaine withdrawal-associated phenotypes and attenuates drug-seeking behavior in rats (7). Building on these early findings, the extent to which intermediary metabolites are directly incorporated as histone modifications and their precise role in cells and organisms remain to be fully elucidated.

Recently, an unanticipated model has arisen for direct interplay between metabolic and epigenetic enzymes (see the figure). Contrary to prevailing dogma that places them exclusively in the mitochondria or cytoplasm, metabolic enzymes can translocate to the nucleus and be recruited to chromatin to regulate local metabolite concentrations in the nucleus or even at specific gene loci. These nuclear metabolic enzymes influence gene expression programs through association with DNA-bound transcription factors, chromatin remodelers, and histone modifiers. Prominent examples are nuclear meta-

bolic enzymes that synthesize acetyl-CoA: acetyl-CoA synthetase (ACSS2) (8, 9) and ATP-citrate lyase (ACLY) (10). By modulating local acetyl-CoA pools, these nuclear metabolic enzymes regulate histone acetylation. Another example is chromatin-bound fumarase, which affects histone methylation by inhibiting 2-oxoglutarate-dependent lysine demethylases (11).

These paradigm-shifting mechanisms have profound relevance for health and disease. Direct metabolic-epigenetic interactions regulate autophagy (9) and cell differentiation (8, 10) and thus can alter the behavior and identity of specific cells. For example, ACLY dynamically regulates histone acetylation in response to growth factor-induced nutrient uptake, thus coordinating cell growth and differentiation with metabolic state (10). These changes, in turn, can also lead to impaired organismal function and disease. In many cases, dysregulation of these pathways is linked to cancer. ACSS2-mediated histone acetylation is implicated in brain tumorigenesis, and ACSS2 concentrations in the nucleus correlate with grades of glioma malignancy in humans (9). Chromatin-bound fumarase promotes tumor growth under glucose deficiency (11).

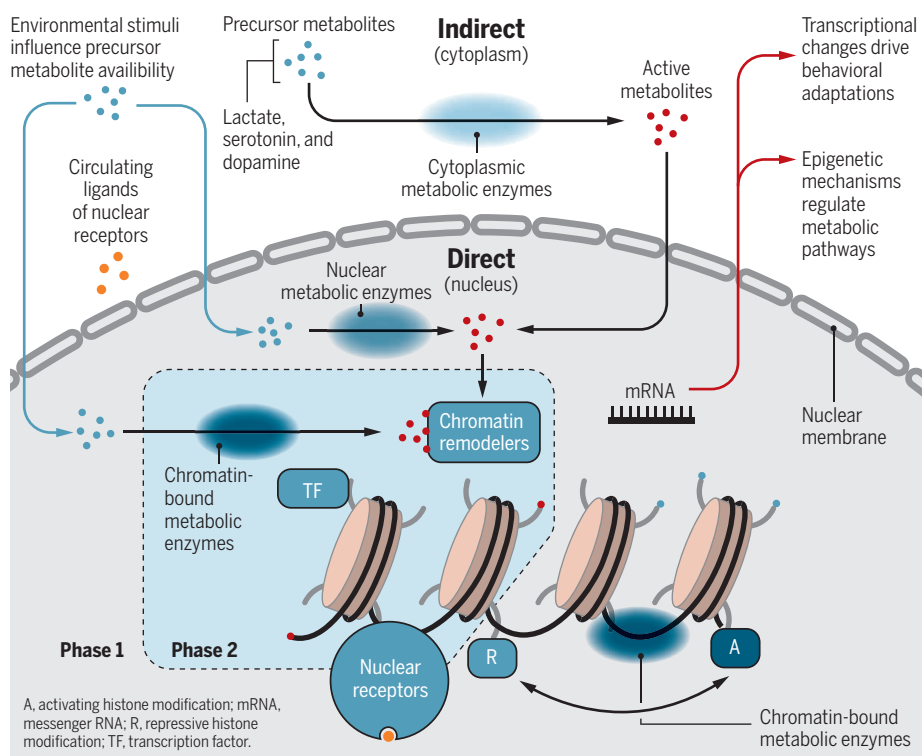
Of particular interest is a nuclear metabolic-epigenetic interface in neurons, where metabolic regulation of histone acetylation serves a key role in spatial learning and memory. Chromatin-bound ACSS2 is enriched at promoters of immediate early genes (IEGs), a set of rapidly induced genes (mostly encoding transcription factors) that drive activity-dependent functional and structural changes in neurons and underlie synaptic plasticity and learning. Reduction of ACSS2 in the hippocampus attenuates long-term spatial memory formation and reconsolidation (8). Further, environmental metabolic fluctuations can influence this pathway. Circulating acetate derived from consumed alcohol is captured by ACSS2 in neuronal nuclei and turned into a local pool of acetyl-CoA. This is incorporated into histone acetylation at IEGs, resulting in the activation of transcription programs that underlie associative spatial memory formation (12). Through this pathway, direct incorporation of alcohol metabolites into brain histone acetylation may play a critical role in encoding drug-related environmental stimuli, which contribute to craving, relapse, and development of substance use disorders.

Future research will expand our understanding of mechanisms that control this metabolic-epigenetic axis. Precise regulation is required to restrict consequences of natural hormonal and metabolic fluctuations that would otherwise change the epigenome, leading to stochastic alterations in transcription. One key regulatory mechanism is the shuttling of metabolic enzymes between the cytoplasm and nucleus. This is gated by phosphorylation or other modifications (9, 11), perhaps to regulate interaction with nuclear import factors and epigenetic enzymes. For example, a recent study demonstrates interactions between the folate pathway enzyme methylenetetrahydrofolate dehydrogenase 1 (MTHFD1) and an acetyl-group reader protein, bromodomain-containing 4 (BRD4). This interaction drives recruitment of MTHFD1 to chromatin to regulate the nuclear concentration of one-carbon (methyl group) metabolites required for transcription, thereby driving cancer cell proliferation (13). Thus, there are likely to be numerous mechanisms to fine-tune metabolic-epigenetic signaling in specific tissues—including distinct neuronal circuits—in response to environmental stimuli, which is critical for effective regulation of behavioral adaptations.

An interesting functional implication of nuclear and chromatin-bound metabolic enzymes is their availability to recycle posttranslational modifications released from histones. This could accomplish dynamic and rapid gene activation. Intriguingly, under oxygen and nutrient limitation, nuclear ACSS2 maintains histone acetylation and gene expression

The metabolic-epigenetic axis and behavior

Through direct and indirect metabolic-epigenetic interactions, transient environmental perturbations can influence gene expression programs, leading to long-lasting functional and behavioral adaptations. For example, chromatin-bound metabolic enzymes can facilitate the direct transfer of metabolites between histone residues and increase the local concentration of active metabolite pools in phase-separated chromatin domains.



by converting acetate released by histone deacetylases into acetyl-CoA, which is then used by histone acetyltransferases (14). Local recapturing of acetate and other metabolites may be essential to maintain or activate new programs of gene expression in nutrient-poor environments. Certain metabolites have profoundly distinct effects on transcription, depending on the histone residue. For example, methyl groups correlate with active [e.g., trimethylated histone H3 lysine 4 (H3K4me)] or repressed (e.g., H3K27me) gene expression. A potential scenario emerges whereby chromatin-bound metabolic enzymes associate with epigenetic enzymes to enable direct transfer of molecules from repressive to activating histone residues (or the reverse) to facilitate rapid transcriptional changes required for cellular and behavioral adaptations.

Another possibility is that nuclear metabolic enzymes are required to maintain metabolite concentrations in phase-separated chromatin domains (15). Free diffusion of metabolites between liquid phases may be limited, and phase-separated chromatin modifiers might thus be largely segregated from the nuclear and cytoplasmic pools of their substrates. Further, chromatin phase separation is regulated by posttranslational histone modifications, for example, acetylation (15), which is under metabolic control. It remains to be seen whether metabolic enzymes phase-separate with epigenetic regulators and whether colocalization to specific phases is required to maintain specific epigenetic signatures.

The dialogue between metabolism and epigenetics within the nucleus presents a promising target for future therapeutic interventions. Certain metabolic-epigenetic pathways could be targeted to inhibit tumor formation and growth (9, 13) or to attenuate certain forms of unwanted memories (e.g., environmental cues that trigger relapse in substance-use disorders or flashbacks in posttraumatic stress disorder) (12). Hence, continued investigation of the metabolic-epigenetic axis holds an exciting potential to launch future therapies. ■

REFERENCES AND NOTES

1. X. Li, G. Egervari, Y. Wang, S. L. Berger, Z. Lu, *Nat. Rev. Mol. Cell Biol.* **19**, 563 (2018).
2. R. M. Evans, D. J. Mangelsdorf, *Cell* **157**, 255 (2014).
3. B. M. Nugent *et al.*, *Nat. Neurosci.* **18**, 690 (2015).
4. K. M. Glastad *et al.*, *Mol. Cell* **77**, 338 (2020).
5. D. Zhang *et al.*, *Nature* **574**, 575 (2019).
6. L. A. Farrelly *et al.*, *Nature* **567**, 535 (2019).
7. A. E. Lepack *et al.*, *Science* **368**, 197 (2020).
8. P. Mews *et al.*, *Nature* **546**, 381 (2017).
9. X. Li *et al.*, *Mol. Cell* **66**, 684 (2017).
10. K. E. Wellen *et al.*, *Science* **324**, 1076 (2009).
11. T. Wang *et al.*, *Nat. Cell Biol.* **19**, 833 (2017).
12. P. Mews *et al.*, *Nature* **574**, 717 (2019).
13. S. Sdelci *et al.*, *Nat. Genet.* **51**, 990 (2019).
14. V. Bulusu *et al.*, *Cell Rep.* **18**, 647 (2017).
15. B. A. Gibson *et al.*, *Cell* **179**, 470 (2019).

10.1126/science.abb4367

PALEOCLIMATE

Steering iceberg armadas

The Asian-Pacific tropics likely instigated millennial-scale climate changes

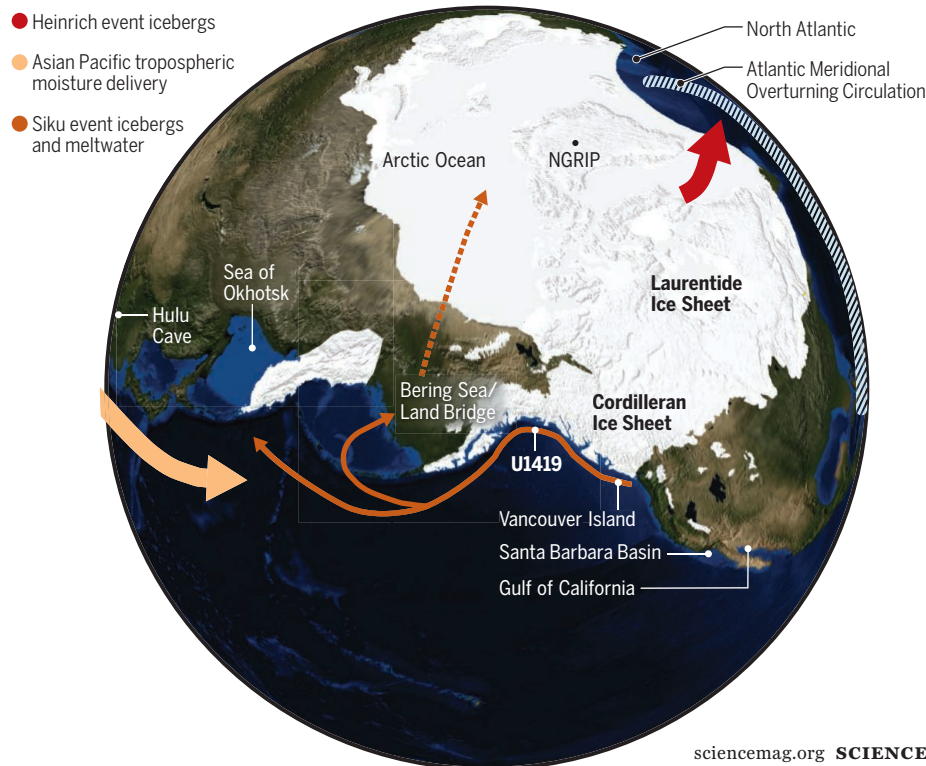
By John M. Jaeger¹ and Amelia E. Shevenell²

Science is theoretically objective, but biases and paradigms often originate from something as fundamental as field site accessibility, data density, or publication date. Such biases may be at the heart of an enduring paradigm in paleoceanography and paleoclimatology—that changes in cold dense water production in the North Atlantic Ocean forced millennial-scale (<1000 year) global climate changes during the past ~50,000 years. These changes were first identified in Greenland ice cores >30 years ago (1). On page 716 of this issue, Walczak *et al.* (2) present compelling data from marine sediments collected in the Northeast Pacific Ocean that add to evidence indicating that the engine of Earth's climate system lies outside the North Atlantic.

The paradigm holds that the intensity of North Atlantic Deep Water (NADW) production controls the rate of the Atlantic Meridional Overturning Circulation (AMOC), a central component of Earth's climate system (3). A key supporting observation is the tenuous temporal link between abrupt North Atlantic climate transitions during the last glaciation, inferred from Greenland ice cores, and melting of marine-terminating ice sheets encircling the North Atlantic (4). Marine sediments deposited in the glacial North Atlantic contain discrete layers of detrital carbonate-rich sands in muddy sediments (5). These coarse deposits, called Heinrich events, suggest melt-out from debris-rich icebergs calved from North America's Laurentide Ice Sheet. Iceberg "armadas" and associated low-salinity glacial meltwater likely disrupted NADW production, thereby slowing

Unraveling a chain of regional climate events 50,000 years ago

Paleoclimate records from places like the Hulu Cave in China record strong Asian monsoon periods preceding Cordilleran melting events by a few millennia. Recurrent changes in iceberg discharges from the Cordilleran Ice Sheet coincide with Sea of Okhotsk sinking water production and precede regional climate changes in Greenland found in North Greenland Ice Core Project (NGRIP) ice cores and produced Heinrich event icebergs. U1419, Integrated Ocean Drilling Program (IODP) Site U1419.



the AMOC and altering global ocean circulation (4). This cascade of events forms the working hypothesis for abrupt (<decade) transitions between cold and warm climate states on millennial time scales during the last glaciation (4, 5). Globally distributed paleoclimate records reveal similar abrupt transitions and millennial-scale events over the same period, and these signals are often linked to North Atlantic processes, even though the root cause of the abrupt events is debated (1, 3–10).

Walczak *et al.* challenge this paradigm, presenting a high-temporal resolution marine sediment record from the Northeast Pacific Ocean (see the figure). They use similar sedimentary archives and data from North Atlantic studies to document changes in regional Northeast Pacific climate. They suggest that similar armadas of icebergs, “Siku events,” calved episodically from the marine-terminating Cordilleran Ice Sheet, when regional atmospheric temperatures were relatively warm. As in the North Atlantic, these Siku events were accompanied by fresh glacial meltwater that restricted vertical ocean mixing, resulting in the apparent aging of intermediate-depth ocean waters (2). The record of Walczak *et al.* indicates that climate signals in the North Atlantic consistently follow, not lead, those observed in the Northeast Pacific, questioning whether NADW perturbations initiate these global events (3–6).

Modern North Pacific climate is influenced by atmosphere-ocean coupling in the tropical Indian and western Pacific Oceans (e.g., El-Niño Southern Oscillation, monsoons) (11). Walczak *et al.* propose that processes in the Asian tropics initiated glacial atmospheric perturbations, which affected ocean circulation and CO₂ cycling on millennial time scales. The authors link glacial Northeast Pacific atmospheric warming and surface ocean freshening to times when the Asian monsoons were relatively strong, as indicated by independently dated Chinese speleothem records.

Critically, the atmospheric and oceanic changes triggering Siku events observed in Walczak *et al.*’s independently dated Northeast Pacific record follow Asian monsoon changes by 1000 to 3000 years and precede North Atlantic Heinrich events and Antarctic and global atmospheric CO₂

changes by 1000 to 1500 years. This pattern of leads and lags supports the hypothesis that the North Pacific is not a passive recipient of North Atlantic climate changes, but rather is close to the source of millennial-scale climate perturbations.

Millennial-scale climate changes attributed to Northern Hemisphere atmospheric teleconnections also are observed in the Santa Barbara basin (7), the Gulf of California (8), and off Vancouver Island (9). However, these records lacked the resolution or detailed independent chronologies to definitively determine the phasing of tropical, North Pacific, and North Atlantic changes. Walczak *et al.* have the necessary archival resolution and chronological control to confidently assess this phasing. The authors’

data provide evidence that globally recognized millennial-scale climate changes likely originated in the tropical Asian-Pacific region. Changes in heat and moisture transport via the westerly Northern Hemisphere jet stream likely explain synchronous rapid Northern Hemisphere warming during the last glaciation (10).

The atmospheric warming melted the Cordilleran Ice Sheet, producing glacial meltwater that may have provided freshwater (via the Bering Sea and Arctic Ocean) that slowed the AMOC (9, 12). Additional evidence from the circum-Antarctic Southern Ocean indicates that overturning driven by tropical atmospheric perturbations regulated glacial-interglacial CO₂ concentrations, rather than the AMOC (13). Taken together, paleoclimate and paleoceanographic records far from the North Atlantic argue that it is time to revisit this paradigm central to paleoceanography and paleoclimate studies since the first high-resolution ice core was drilled in Greenland (1, 3). ■

REFERENCES AND NOTES

1. W. Dansgaard *et al.*, *Nature* **364**, 218 (1993).
2. M. H. Walczak *et al.*, *Science* **370**, 716 (2020).
3. A. Ganopolski, S. Rahmstorf, *Nature* **409**, 153 (2001).
4. W. S. Broecker, *Nature* **372**, 421 (1994).
5. G. C. Bond, R. Lotti, *Nature* **267**, 1005 (1995).
6. N. Brown, E. D. Galbraith, *Clim. Past* **12**, 1663 (2016).
7. I. L. Hendy, J. P. Kennett, *Geology* **27**, 291 (1999).
8. T. Bhattacharya *et al.*, *Nat. Geosci.* **11**, 848 (2018).
9. I. L. Hendy, T. Cosma, *Paleoceanography* **23**, PA2101 (2008).
10. S. K. Praetorius, A. C. Mix, *Science* **345**, 444 (2014).
11. J. Emile-Geay, *J. Geophys. Res.* **108** (C6), 3178 (2003).
12. S. K. Praetorius *et al.*, *Sci. Adv.* **6**, eaay2915 (2020).
13. A. C. Mix *et al.*, in *Mechanisms of Global Climate Change at Millennial Time Scales*, P. U. Clark *et al.*, Eds. (Geophys. Monogr. Ser., American Geophysical Union), vol. 112, pp. 127–148.

10.1126/science.abe8461

CORONAVIRUS

Public health during the pandemic in India

Data from two Indian states give detail about variation in transmission and disease outcomes

By Jacob John¹ and Gagandeep Kang²

In March 2020, with coronavirus disease 2019 (COVID-19) threatening to overwhelm India’s fragile health care ecosystem, the country combined a stringent lockdown of its 1.37 billion population with a program of surveillance and containment of varied effectiveness across states. Testing and data management systems were set up, but the paucity of publicly available data, especially in the initial phase of the pandemic, limited understanding of disease epidemiology and transmission dynamics as well as the effectiveness of control measures. On page 691 of this issue, Laxminarayan *et al.* (1) present findings from government-implemented surveillance during the first 4 months of the pandemic in the two southern states of Tamil Nadu and Andhra Pradesh in India. They use data from, and after, the lockdown period to make important observations on the dynamics of infection, transmission, and risk factors. This collaboration between the state governments of Tamil Nadu and Andhra Pradesh and academic researchers is a valuable template for federal government agencies.

The public health response to infectious disease outbreaks is founded on the ability to mount a coordinated strategy that combines measuring and tracking cases to assess the efficacy of interventions. For COVID-19, testing, tracing, treating and isolating cases, quarantining contacts, as well as widespread mask wearing and social distancing were, and are, the tools for transmission control. India’s population is second only to China’s

¹Department of Community Health, Christian Medical College, Vellore TN, India. ²Division of Gastrointestinal Sciences, Christian Medical College, Vellore TN, India. Email: gkang@cmcvellore.ac.in

¹Department of Geological Sciences, University of Florida, Gainesville FL 32611-2120, USA. ²College of Marine Science, University of South Florida, St. Petersburg, FL 33701, USA. Email: jmjaeger@ufl.edu; ashevenell@usf.edu

approximately 1.41 billion, in one-third of the land area, and so when severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) began its relentless spread outside China, there were concerns about the ability of low- and middle-income countries (LMICs) to deploy public health tools, given their limited health care capacity.

Although there were limitations in tracking in many parts of the world, the data from hospitals and laboratories were widely available, allowing modelers in academia and public health systems to use data from China and the industrialized world early in the pandemic to build predictive models based on the government strategies that were developed (2). Although useful, these models are sensitive to underlying assumptions about contact patterns and transmissibility of the infection, which are affected by population density, occupation, and social structures. In India and other LMICs, the lack of information sharing for analysis between government agencies with access to the data and academic groups with the skills to perform these analyses resulted in control strategies being widely debated because the evidence base for policy was unclear (3, 4).

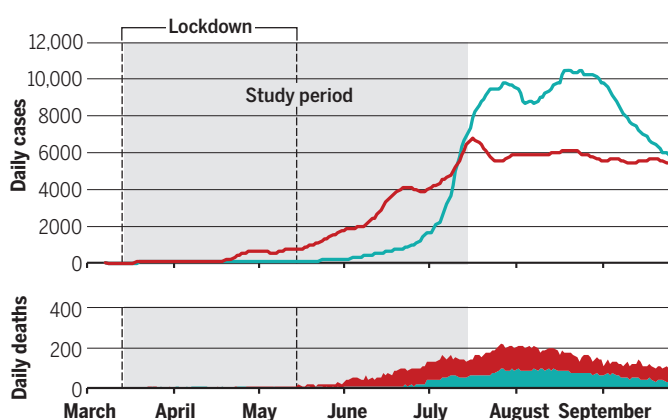
Despite the perpetual underinvestment in public health in India, both Andhra Pradesh and Tamil Nadu have functional public health departments. Laxminarayan *et al.* describe scaling of contact tracing to reach more than 3 million exposed contacts and collect epidemiological and laboratory data from 575,051 contacts of cases in the first 4 months of the pandemic, when such extensive contact tracing was most likely to be beneficial. The surveillance was not perfect. There were variations in the effectiveness of screening across districts, with most cases having an improbably low number of exposed contacts. Testing strategies changed multiple times as the public health response adapted to emerging challenges, complicating analyses of infection rates and infection fatality ratios. Despite these limitations and the subsequent escalation in cases (see the figure), the authors have generated important insights.

The study highlights the importance of “superspreaders” in fueling the pandemic. Although there were limitations in contact tracing, and hence in the evaluation of secondary cases and of transmission chains, the finding that a minority of cases were associated with most transmission events (superspreading) is consistent with the emerging

data across the globe (5, 6). The determinants of superspreading are not well understood, but it is likely to be a function of social interaction patterns of the infectious host, the environment, and biological characteristics of the infectious agent. Identifying the characteristics of settings where superspreading events are likely will help target control measures and screening for those settings to maximize limitation of transmission. However, focused control measures must be balanced against the potential for further stigmatization of individuals with COVID-19 beyond that already reported (7).

Cases and deaths in two southern states of India

Trends in case detection (7-day moving average is shown) and coronavirus disease 2019 (COVID-19) fatalities in Andhra Pradesh (●) and Tamil Nadu (●) highlight low case fatality and the effects of early interventions.



The contact tracing data suggests two important findings regarding transmission. First, there is an increased likelihood of cases infecting contacts of similar ages, and second, that children under 14 are active participants in transmission. These findings probably reflect prevalent social mixing patterns in India (8) but raise concerns about the possibility of enhanced transmission when poorly ventilated, crowded schools reopen (9). Children have been implicated in the transmission of other respiratory infections to the elderly when intergenerational contact is frequent (10). Careful monitoring of transmission by children will be vital to shield vulnerable older adults and those with comorbidities. At the other end of the age spectrum, the low mortality rate in the elderly as reported by Laxminarayan *et al.* indicates the need to better understand the composition of older age groups by socioeconomic status and comorbidities to clarify the contribution of the survivor effect. The lower death rates may also reflect that those who live longer in LMICs are more healthy than people of similar ages in high-income countries (11).

More than half the reported deaths in these two states occurred within 6 days of diagnosis compared with 16 days in China (12). There are several potential explanations. During the lockdown and with concerns about transmission, individuals who were ill may have come to health care facilities late in their illness. It is also possible that owing to limited capacity, testing was reserved for those who had more advanced symptoms and were potentially more likely to die. In the early phase of the pandemic, the treatment facilities may have been inadequately prepared, equipped, and staffed. The 26%

improvement in mortality risk in recent months in these states (1) may partly be attributed to improved access to testing, treatment, and health care infrastructure over time.

The observation that traveling in proximity to a case for a long duration was a risk factor for transmission merits special consideration for India, where most interstate transport is by train, bus, and truck. Studies in other parts of the world on air travel have also documented spread (13). Most of these studies report data before widespread mask use, and experimental and real-world data obtained from using different modes of transport with well-described air circulation must be used to inform travel guidance (14). More studies from

diverse settings at this stage of the pandemic, particularly including data from serosurveys, are needed to better understand factors determining transmission, plan interventions, and predict outcomes in LMICs. ■

REFERENCES AND NOTES

1. R. Laxminarayan *et al.*, *Science* **370**, eabd7672 (2020).
2. S. Flaxman *et al.*, Imperial College COVID-19 Response Team, *Nature* **584**, 257 (2020).
3. The Lancet, *Lancet* **396**, 867 (2020).
4. Indian Public Health Association, Indian Association of Epidemiologists, Indian Association of Social and Preventive Medicine, “3rd joint statement on covid-19 pandemic in India—Public health approach for covid-19 control” (2020); <https://bit.ly/2ITbky8>.
5. A. Endo, S. Abbott, A. J. Kucharski, S. Funk, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, *Wellcome Open Res.* **5**, 67 (2020).
6. M. S. Y. Lau *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 22430 (2020).
7. E. Cave, *Asian Bioeth. Rev.* **1** (2020). 10.1007/s41649-020-00118-2
8. S. Kumar *et al.*, *PLOS ONE* **13**, e0209039 (2018).
9. J. P. Majra, A. Gur, *J. Glob. Infect. Dis.* **2**, 109 (2010).
10. S. Cauchemez, A. J. Valleron, P. Y. Boëlle, A. Flahault, N. M. Ferguson, *Nature* **452**, 750 (2008).
11. S. Afshar, P. J. Roderick, P. Kowal, B. D. Dimitrov, A. G. Hill, *BMC Public Health* **15**, 776 (2015).
12. T. Chen *et al.*, *BMJ* **368**, m1091 (2020).
13. J. Chen *et al.*, *Travel Med. Infect. Dis.* **36**, 101816 (2020).
14. C. O. Buckee *et al.*, *Science* **368**, 145 (2020).

10.1126/science.abe9707



CONSERVATION

Species protection will take more than rule reversal

Key improvements are needed for implementation of the Endangered Species Act

By Ya-Wei Li¹, Joe Roman², David S. Wilcove³, Timothy Male¹, Holly Doremus⁴

Species are disappearing at an alarming rate, with global estimates of about a million species facing extinction (1). The Endangered Species Act (ESA), the primary—and often only—means in the United States to prevent extinctions, is justly celebrated as perhaps the strongest model for endangered species protection worldwide. Since its adoption, however, the ESA has faced unabated controversy because it can restrict economic activities and because its implementation often appears inconsistent. With the explicit goal of reducing “unnecessary regulatory burdens,” the Trump administration in 2019 finalized the most comprehensive changes in more than two decades to the regulations that implement the ESA (2). Some of the changes will make it harder to protect species and their habitats; none will directly further the Act’s goal of recovering species. For example, the changes limit the government’s ability to protect habitat that species need to adapt to climate change (3) and make it harder for the public to hold the federal government accountable for activities that further imperil species (4). Opposition to the changes was swift and

ardent among many environmentalists, scientists, and the public. Opposition to the administration’s changes, however, should not overshadow the need for improvements to how the ESA is administered to make it more effective. Simply revoking recent changes will not solve these underlying problems.

Without deeper reforms to address underlying problems, implementation of the ESA by the U.S. Fish and Wildlife Service (FWS) and the National Marine Fisheries Service (NMFS) (“the Services”) will remain ad hoc and insufficiently explained (5). This ambiguity invites political intervention that undercuts species protection and public confidence in ESA decisions, triggers litigation that is costly for all parties, and polarizes the law. Finding solutions to these problems could lead to broad bipartisan initiatives to stem biodiversity loss and to increase funding for the ESA by reauthorizing the law.

DEFINING THE “FORESEEABLE FUTURE”

One of the controversial revisions pertains to how the Services determine the “foreseeable future,” which is used to decide whether species merit listing as “threatened” under the ESA. The law recognizes two levels of threat: Species may be “endangered”—that is, presently in danger of extinction—or

The 2017 decision to not protect the Pacific walrus (shown) was based on projections to only 2060, despite a 2012 decision to protect the Arctic ringed seal that looked out to 2100.

“threatened,” which means likely to become endangered in the “foreseeable future.” In 2009, the Services first articulated their understanding of the term “foreseeable future,” declaring that it covers the time frame over which predictions of the extent of threats and their impact on species are “reliable” (6). The new regulations provide that the foreseeable future “extends only so far into the future as the Services can reasonably determine that both the future threats and the species’ responses to those threats are likely.” The agencies’ explanation “clarifies” that “likely” means “more likely than not.” Thus, whether this new definition will change established practice turns on the difference, if any, between whether predictions of the future are “reliable” or “likely.” The Services claim that there is no difference, whereas many environmentalists see an intent to ignore climate change impacts on species (7).

Whatever the linguistic change means, the underlying problem of inconsistent and inadequately explained treatment of the foreseeable future remains. The Services have applied notably different interpretations to species facing similar threats. When NMFS listed the Arctic ringed seal (*Phoca hispida hispida*) in 2012, for example, it estimated the threat of reduced sea ice and snow cover out to the year 2100, stating that it was able to “reliably” forecast ~90 years into the future on the basis of models of how global greenhouse gas levels would affect the Arctic environment (8). But when FWS declined to list the Pacific walrus (*Odobenus rosmarus divergens*) in 2017 in the face of similar threats, it limited its evaluation to 2060 because it considered any conclusions beyond that date to be “based on speculation, rather than reliable prediction” (9). The Services did not articulate any difference in the natural histories of the seal or walrus that could justify this difference. The state of Alaska has petitioned NMFS to remove the Arctic ringed seal from the endangered species list in part on the basis of this discrepancy (10).

Courts and researchers also have expressed concerns about inconsistencies or arbitrariness in how FWS has interpreted the “foreseeable future” (11, 12). For example, when a court rejected FWS’s listing of the

¹Environmental Policy Innovation Center, Washington, DC, USA. ²Gund Institute for Environment, University of Vermont, Burlington, VT, USA. ³Princeton School of Public and International Affairs and Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, USA. ⁴School of Law, University of California, Berkeley, Berkeley, CA, USA. Email: jake@policyinnovation.org

northern long-eared bat (*Myotis septentrionalis*) as threatened instead of endangered, the agency did not dispute that its evaluation of threats, which extended only 8 to 13 years into the future, was an irrational approach to interpreting the foreseeable future (13). The new foreseeable future definition does not fix these problems, but neither will restoring the prior one.

What is required is consistency and transparency. That will come only if the Services issue guidance that will both hold them accountable and explain the principles that motivate their decisions. Such guidance should ensure, for example, that projections about both the geophysical aspects of climate change and their effects on species are consistent across comparable situations. Once NMFS concluded that the extent of sea ice loss was reliably foreseeable to 2100, any conflicting decisions should explain why that conclusion was wrong or why it merits revision in light of new data. Given that 2100 is embedded in many of the global climate projections constructed by the Intergovernmental Panel on Climate Change, we suggest this date as a conservative starting point for assessing species vulnerable to climate change. Although climate change could affect individual species differently, the Services should clearly explain the evidence for these varied effects and the justification for differential treatment.

EXPLAINING DISCRETION

The key protections of section 9 of the ESA apply only to endangered species, not threatened species. Protections include restrictions on importing endangered species into the United States, trafficking in them or their parts, and harming or harassing endangered animal species by other means, including habitat destruction. For threatened species, Congress gave the Services the authority to decide on a species-by-species basis which protections to apply. FWS has long extended by default the full protections of section 9 to all threatened species while retaining discretion to modify those protections on a species-by-species basis through a special rule issued under section 4(d) of the ESA. The recent regulatory revisions withdrew those default protections for future listings, requiring FWS to issue a 4(d) rule whenever it seeks to extend any protection to those species and aligning the agency's approach with that of NMFS, which has never extended default protection to all threatened species. Despite this reversal in FWS policy, the agency is still able to offer threatened species as many or as few protections as it deems necessary for conservation—as has always been the case.

The problem is that the Services have never issued clear guidance on how they will exercise this discretion, nor have they ad-

equately explained their choices. Under the ESA, the Services “may” offer threatened species none, some, or all of the section 9 protections. In the context of agricultural activities, for example, FWS offered the Gunnison sage-grouse (*Centrocercus minimus*) full protections. By contrast, the 4(d) rule for the related lesser prairie chicken (*Tympanuchus pallidicinctus*) exempted all routine agriculture on cropland maintained in cultivation (14). The agency may have had valid reasons for this discrepancy, but they have never publicly explained those reasons. Exemptions in 4(d) rules thus often appear as ad hoc decisions influenced by political pressure to minimize regulatory impacts of listing a species. This concern can trigger litigation from conservation groups, resulting in further controversy and expenditure on lawsuits.

The Services should develop policy that resolves key issues pertaining to protection of threatened species through 4(d) rules. That policy should directly address the question that they have long ducked: What protections meet the ESA's standard of being “necessary and advisable” to conserve threatened species? Having a policy that states the relevant principles would limit the Services' tendency to bow to political pressures, creating an ad hoc patchwork of protections. It could also help assure landowners that voluntary efforts at conservation will not bring a heavy regulatory crackdown. At a minimum, activities that would undercut a species' recovery should be regulated through 4(d) rules, and activities that promote recovery should be strongly considered for exemption. An example is the recent 4(d) rule for the Louisiana pine snake (*Pituophis ruthveni*), which exempts forestry activities that improve the snake's habitat but regulates intensive mechanical forestry practices that can degrade that habitat (15). Second, the Services should commit to finalizing the protections that a threatened species needs when it is listed, unless there is substantial uncertainty about whether the protections will benefit the species. By addressing these and other basic issues, the agencies can help ensure that protections for threatened species are adequate and predictable.

NEW APPROACHES

The improvements above focus on issues that are addressed in the recent rulemaking and that can be addressed without legislation, but other reforms also deserve priority. For example, in 2016 FWS developed a plan to address its backlog of decisions on whether to list hundreds of species under the ESA. FWS will need to diligently implement the plan to reduce litigation over delayed listing decisions—something it has so far failed to do, partly because of political intervention.

A new regulatory and funding package for working with private landowners to conserve imperiled species, including dedicated staffing for ESA voluntary conservation initiatives and tax benefits for easements and donations of private land for rare-species conservation, would unlock recovery opportunities for many species that rely on private lands. Conservation on federal lands could benefit from legal incentives for federal agencies to carry out actions that go beyond the minimum required by the ESA, such as rewarding agencies with greater management flexibility when they help a species exceed its recovery milestones. A new wildlife data and technology initiative could bring ESA implementation into the 21st century by taking advantage of open and machine-readable data, remote sensing data, and other technological innovations to help monitor species and their habitats. Such advances offer some of the best opportunities to understand how climate change will affect the nearly 2400 species protected by the ESA.

To keep pace with our biodiversity crisis, the ESA will need to go well beyond the status quo. Let the current controversy over the revised regulations serve as the starting point to finding meaningful solutions and having deeper discussions of what must be done to conserve imperiled species in the United States and elsewhere. The passage of the Great American Outdoors Act reminds us that conservation can still be a bipartisan issue. The reforms we suggest could help bring us closer to consensus on the ESA. ■

REFERENCES AND NOTES

1. Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services, “Global assessment report on biodiversity and ecosystem services” (2019).
2. FWS, NMFS, “Trump administration improves the implementing regulations of the Endangered Species Act” (2019); www.fws.gov/news/ShowNews.cfm?ref=trump-administration-improves-the-implementing-regulations-of-the-ESA_ID=36443.
3. FWS, NMFS, 84 Fed. Regist. 84,45020 (2019) [to be codified at 50 Code of Federal Regulations (C.F.R.) § 424.12] (2019).
4. FWS, NMFS, Fed. Regist. 84,44976 (2019).
5. H. Doremus, *Wash. Univ. Law Q.* 75, 1029 (1997).
6. U.S. Department of the Interior, “The Meaning of ‘Foreseeable Future’ in Section 3(20) of the Endangered Species Act,” memorandum M-37021 (2009).
7. R. Beitsch, *The Hill*, 12 August 2019; <https://thehill.com/policy/energy-environment/457086-trump-administration-rolls-back-endangered-species-protections>.
8. NMFS, Fed. Regist. 77,76706 (2012).
9. FWS, Fed. Regist. 82,46618 (2017).
10. State of Alaska et al., “Petition to Delist the Arctic Subspecies of Ringed Seal (*Phoca hispida hispida*) Under the Endangered Species Act” (2019).
11. *Defenders of Wildlife v. Jewell*, 176 F. Supp. 3d 975 (D. Mont. 2016).
12. J. D'Elia, S. McCarthy, *Bioscience* 60, 751 (2010).
13. *Center for Biological Diversity v. Everson*, no. 15-CV-477, 2020 WL 437289 (D.D.C. 28 January 2020).
14. FWS, Fed. Regist. 79,20074 (2014).
15. FWS, Fed. Regist. 85,11297 (2020).

ACKNOWLEDGMENTS

We thank M. Bean for feedback on a draft of this manuscript.

10.1126/science.abb3806



INFORMATION SCIENCE

The ascent of Wikipedia

Scholars reflect on 20 years of crowdsourced knowledge

By **Andrew Robinson**

In 2005—not long after the founding of Wikipedia by Jimmy Wales and Larry Sanger in early 2001—academic experts commissioned to compare 42 articles published in *Encyclopaedia Britannica* and Wikipedia relating to science found an average of three errors in the *Britannica* entries and four in Wikipedia, suggesting a comparable level of accuracy (1). Yet in 2007, Michael Gorman, former president of the American Library Association, argued scornfully that “A professor who encourages the use of Wikipedia is the intellectual equivalent of a dietician who recommends a steady diet of Big Macs with everything” (2). Gorman’s article reflected the widespread skepticism at the time about the reliability of an encyclopedia that anyone can edit.

Today, Wikipedia is the world’s leading encyclopedia. Every month, 1.5 billion unique devices worldwide access it 15 billion times, with more than 6000 page views per second. Meanwhile, *Encyclopaedia Britannica*—last printed in 2010—is now “all but dead” online, according to scholar Heather Ford in her essay in *Wikipedia @ 20*.

The reviewer is the author of *The Last Man Who Knew Everything* (Oneworld Publications, 2007) and *Genius: A Very Short Introduction* (Oxford Univ. Press, 2011). Email: andrew@andrew-robinson.org

The book’s 22 essays are wide-ranging, often intellectually engaging, and, in parts, stylishly written. Its 34 contributors include, fittingly, academics and nonacademics based in many countries, although predominantly in the United States. Its U.S.-based editors, Joseph Reagle and Jackie Koerner, are (respectively) a professor of communication studies and a qualitative research analyst for online communities who also acts as the community health consultant for the Wikimedia community.

Wales neither contributes an essay nor makes much of an appearance in the book, although he does provide a pithy blurb saluting the “hard-won wisdom of its contributors, the novel reflections of scholars, and the necessary provocations of those working to shape [Wikipedia’s] next twenty years.” These provocations are examined in detail and range from what to do about microaggressive community editing of novice Wikipedia contributors to how Wikimedia, the nonprofit organization that hosts Wikipedia, might address the lack of reliable internet access experienced by four billion people worldwide.

Perhaps the most important internal controversy discussed is the bias of Wikipedia’s contributors and entries. The English Wikiped-

Surpassing the expectations of its early critics, Wikipedia is now the world’s leading encyclopedia.

ia has more than 31,000 active contributors, yet only 11 language editions have more than a thousand active contributors, and more than half of the editions have fewer than 10. Moreover, a recent Wikimedia Foundation survey revealed that about 91% of Wikipedia editors are male and 77% are white. As of late 2019, only 18% of biographies on Wikipedia were about women. Far fewer were about nonwhites. A number of nonmale and nonwhite contributors to the book describe their own, often unavailing, efforts to improve these statistics.

There is also the tricky issue of determining the appropriate relationship that should exist between Wikipedia’s language editions. Each is written independently of the other 284 editions, although “a contributor may consult an existing article in another language edition when writing a new article, or they might even use the content translation tool to help with translating one article to another language, but there is nothing to ensure that articles in different language editions are aligned or kept consistent with each other,” notes Denny Vrandečić, founder of both Wikidata and the Croatian Wikipedia. “This is often regarded as a contribution to knowledge diversity since it allows every language edition to grow independently of all other language editions,” he adds, begging the question: “Would creating a system that aligns the contents more closely with each other sacrifice that diversity?”

Ironically, one weakness of the book—common in edited collections—is the infrequency of cross-referencing between articles. Another is the sweeping summary conclusion by Katherine Maher, executive director of the

Wikimedia Foundation, that “Wikimedia reminds us that the greatest thing we will ever build is the thing which we build with others.” This surely underestimates the advances in human knowledge that have come from solitary thinkers (many of them specialists)—a universally popular part of Wikipedia’s entries—rather than from collaborative groups. That said, anyone interested in the history, current constitution, and possible

future development of a singular contemporary global phenomenon will be stimulated by this anniversary collection. ■

REFERENCES AND NOTES

1. J. Giles, *Nature* **438**, 900 (2005).
2. M. Gorman, “Jabberwiki: The Educational Response, Part II,” *Encyclopaedia Britannica Blog* (2007); <http://blogs.britannica.com/2007/06/jabberwiki-the-educational-response-part-ii/>.

10.1126/science.abf2433



Wikipedia @ 20
Joseph Reagle and
Jackie Koerner, editors
MIT Press, 2020. 376 pp.

SCIENCE LIVES

Recognizing the work of women

Despite several stumbles, a new volume about women in paleontology will likely prove valuable to future scholars

By **Jess Miller-Camp**

The history of vertebrate paleontology, like that of other scientific disciplines, has traditionally been told and shaped through a masculine filter. Annalisa Berta and Susan Turner aim to rectify this bias, and their new book, *Rebels, Scholars, Explorers*, is a stepping stone along that path, synthesizing observations about the past, present, and future of women's contributions in the field. The book is the first major effort to bring the work of several centuries' worth of silenced people to light. Any such endeavor will inevitably have gaps, but it will also serve as a reference for future projects—particularly with regard to premodern individuals, details about whom are often much harder to track down.

Rebels, Scholars, Explorers can perhaps best be thought of as a three-part work. In the first section, the authors offer a broad overview of the often hostile climate within which women in vertebrate paleontology have long operated. In the second, they present short biographical sketches of a sampling of women throughout the field's history. In the third, they provide an assessment of where things currently stand and where we might go from here.

The book includes the contributions of many early modern women I had never heard of. Among them is Maria Pavlova, who, although she was an eminent Russian paleontologist and professor in 1919, could do nothing when a close colleague rejected her highly qualified student (a woman) for a professional position. In her memoir, Pavlova wrote of the incident, "I'm feeling as if I returned from a very hard funeral."

A number of the biographical sketches also contain anecdotal gems. Taphonomist Kay Behrensmeyer's, for example, references her nonchalance about a venomous green mamba at her field site in Cameroon. And Elizabeth (Betsy) Nicholls's entry notes how, at the age of 12, she wrote to renowned natu-

ralist Roy Chapman Andrews to ask if girls could be paleontologists—and how she kept his encouraging reply as a treasured token.

The inclusion of a chapter on modern individuals whose careers are not primarily in original research is refreshing, given academia's general habit of ignoring such contributors. Included here are women such as Akiko Shinya, the chief vertebrate fossil preparator at the Field Museum, and Tsiory Andrianavalona, who leads an outreach center in Madagascar called ExplorerHome



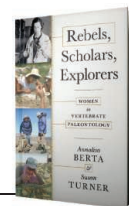
Avian paleontologist Hildegard Howard examines bird bones at the Natural History Museum of Los Angeles County.

that seeks to inspire scientific curiosity in Malagasy students.

It was incredible and heartening to see, throughout the book, so many citation-rich sections exclusively referencing works led by women. The authors' account of vertebrate paleontology in the Soviet Union, for example, discusses the work of a host of women studying Paleozoic fishes. Such passages make stark the extent to which the contributions of women are often overlooked.

The book's interviews contain insightful

Rebels, Scholars, Explorers: Women in Vertebrate Paleontology
Annalisa Berta and Susan Turner
Johns Hopkins University Press, 2020. 344 pp.



observations and advice from women such as Margery Coombs, Catherine Badgley, and Romala Govender, as well as from male paleontologists who have mentored women. However, some interviewees aim for equality rather than equity. Reading such passages was frustrating, but their inclusion accurately reflects the prevalence of misguided efforts at allyship. (This section of the book is presented without commentary and is bound to spark some very good discussions in paleo reading groups.)

As a book about the history of a field whose infancy lies in the West, *Rebels, Scholars, Explorers* has a largely Occidental focus. Discussion of women in other theaters picks up, albeit inconsistently, as the book proceeds forward in time. The authors thoroughly cover the scientific efforts of mid-20th-century Chinese women, for example, but only a single woman from and living in India—Sanjukta Chakravorti, a recent graduate—is included.

The book's final chapter begins to address intersectionality by discussing the additional challenges faced by women paleontologists who are also ethnic minorities. Here, Anusuya Chinsamy-Turan's account of the restrictions placed on her during South Africa's apartheid was particularly poignant. However, the authors' handling of transgender issues, another intersectional identity, is disappointing.

Although they rightly include at least one trans woman in their account, they also include James Robinson, who, prior to his transition, studied plesiosaur anatomy and behavior. One wonders how he would feel about being referred to by his pretransition name—a practice known as deadnaming, which most who transition consider to be disrespectful—and, more generally, about his inclusion in a book about women. It is vital that anyone studying or writing about gender seek out noncisgendered perspectives, lest errors be introduced and individuals harmed.

Despite these shortcomings, *Rebels, Scholars, and Explorers* is a strong backbone upon which future scholarly work can build. I look forward to seeing the rest of the skeleton uncovered. ■

10.1126/science.abe5779

The reviewer is at the Department of Earth and Atmospheric Sciences, Indiana University Bloomington, Bloomington, IN 47405, USA. Email: jessmc@iu.edu



The Oriental stork (*Ciconia boyciana*) is threatened by human activities in its migratory stopover points in China.

Edited by Jennifer Sills

Aquaculture jeopardizes migrating Oriental storks

The Oriental stork (*Ciconia boyciana*), once the most common bird of the Far East, is now listed in Appendix I of the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES) and classified as Endangered on the International Union for Conservation of Nature's (IUCN's) Red List (1, 2). Over the past few decades, illegal poaching, habitat loss, and environmental pollution have caused a decline in the Oriental stork population in its Northeast Asian breeding grounds (3). Fewer than 3000 individuals remained worldwide by 2018 (2). The Oriental stork now faces a severe survival threat at migratory stopover sites in China's Bohai coastal region, where human disturbances drive habitat degradation (4).

The Qilihai and Caofeidian wetland reserves, located north of the Bohai Bay, are the most important stopover sites for the Oriental stork (5). More than 2000 individuals refuel in these wetlands during their fall migration (6). However, large areas of these reserves (including about 80 to 90% of the Caofeidian wetland) have been leased to aquacultural farmers to create fishponds (7). The farmers dislike foraging storks and often try to scare them away with fire-crackers (4). For aquaculture, wetlands are maintained at a relatively high water level, which makes foraging difficult for storks

and prevents them from finding enough food (4). As a result, the storks relocate to private fishponds, where they face an increased risk of poisoning and poaching. In 2019, 19 storks were poisoned in the Qilihai and Caofeidian wetlands (8).

The Oriental stork has been categorized as a terrestrial species under state protection (with a beneficial, economic, or scientific value) since 2000 (9), but the species' extremely limited population size indicates that this designation is not enough. To better protect the Oriental stork, the Chinese government is currently updating the special state protection list and changing the protection of this bird to the highest first-class level (10). In addition, immediate habitat restoration is required. Fishponds in protected areas must be restored to natural wetlands to create favorable living conditions and reverse population decline. Local governments should rapidly formulate eco-compensation measures and publicize wildlife protection to mediate human-bird conflicts. Only by taking action to protect this species and its habitat can we prevent its looming extinction.

Yanfeng Sun^{1,2}, Liqiang Du³, Yuan Yin², Yuefeng Wu², Aijun Cai³, Dongming Li^{2*}

¹Ocean College, Hebei Agricultural University, Qinhuangdao 066000, China. ²Key Laboratory of Animal Physiology, Biochemistry and Molecular Biology of Hebei Province, College of Life Sciences, Hebei Normal University, Shijiazhuang 050024, China. ³College of Marine Resources & Environment, Hebei Normal University of Science & Technology, Qinhuangdao 066600, China.

*Corresponding author.

Email: lidongming@hebtu.edu.cn

REFERENCES AND NOTES

1. CITES, "Appendices" (2020); <https://cites.org/eng/app/appendices.php>.
2. BirdLife International, "*Ciconia boyciana*" (IUCN Red List of Threatened Species, 2018).
3. Y. Yamada *et al.*, *Ecol. Res.* **34**, 277 (2019).
4. IUCN Commission on Environmental, Economic, and Social Policy, "A reflection on protected areas in serving wildlife migration: Endangered Oriental storks" (2019).
5. W. Peng *et al.*, *Int. J. Ecol.* **9**, 108 (2020) [in Chinese].
6. Z. J. Zhao, *Avifauna of China* (Jilin Science and Technology Press, vol. 1, 2001) [in Chinese].
7. "Ninety percent of Caofeidian wetland was outsourced and turned into fishponds" [China Biodiversity Conservation and Green Development Foundation (CBCGDF), 2019].
8. "In-depth analysis of 'murder' on the journey: What's behind the continuous death of Oriental white stork? Is it a 'gang' crime?" (CBCGDF, 2019); www.cbcdgf.org/NewsShow/4854/10817.html [in Chinese].
9. National Forestry and Grassland Administration, Government of China, "Lists of terrestrial wildlife under state protection, which are beneficial or of important economic or scientific value" (2000); www.forestry.gov.cn/main/3954/content-959027.html [in Chinese].
10. National Forestry and Grassland Administration, Government of China, "List of wildlife under special state protection (draft)" (2020); www.forestry.gov.cn/main/153/20200619/092731170435586.html [in Chinese].

10.1126/science.abf2089

Indigenous rights to Patagonia's Guafo island

In September, private investors put 20,000-ha Guafo island up for sale for US \$20 million (1). The island, located in northern Chilean Patagonia, is a stronghold of unique biodiversity (2, 3) and a biocultural heritage site for Patagonia Indigenous groups and the country (4). The owners bought the island a decade

ago with plans to conduct coal mining operations (1), but after Chile made a climate change commitment to become carbon neutral by 2050 (5), the difficulty obtaining mining permits made them rethink their investment. The Chilean government should protect these valuable ecosystems by deeding the island to the Indigenous people who claim it.

The challenges facing Guafo island are emblematic of environmental problems throughout Patagonia. A growing exotic salmon aquaculture industry has brought about environmental degradation (6). Climate change-induced drought has led to harmful algal blooms (6, 7) that may have contributed to the stranding of hundreds of endangered Sei whales (8). Terrestrial ecosystems face increasing tensions from tourist developments, a growing human population, peat bog degradation, exotic species invasion, and climate change (6), which threatens glacial freshwater reserves (9) and increases the likelihood of fires (6).

Disrespecting Indigenous peoples' heritage in Patagonia threatens the region's biodiversity. Degrading the capacity of carbon sequestration in forest soils, peatbogs, and kelp forests (6) and disrupting the carbon sequestration processes fostered by large whales and other marine vertebrates (10) will initiate a perfect storm of increasing warming and ecosystem degradation with global consequences. It is incumbent upon the Chilean government to set an example for Patagonian policy by protecting Guafo island and its surrounding seascapes, requiring that the salmon industry withdraw operations from protected waters, and giving Indigenous people the rights to their ancestral lands. In 2008, Chile passed the Mapuche-Lafkenche Marine and Coastal Areas for Indigenous Peoples (MCAIP) law (11). The legislation enables the allocation and administration of coastal marine areas to Indigenous communities, who can ensure the sustainability and conservation of marine resources and ecosystems. There is already an MCAIP claim for Guafo island

by Indigenous communities from nearby Chiloé island (12), the "Wafo Wapi ancestral land for conservation." The Chilean government should support this claim.

Pablo A. Marquet^{1,2*}, Juan Carlos Castilla¹, Aurora Gaxiola^{1,2}, Rodrigo Huckle-Gaete³, Alfredo Pena-Vega⁴

¹Departamento de Ecología, Facultad de Ciencias Biológicas, Pontificia Universidad Católica de Chile, CP 8331150, Santiago, Chile. ²Instituto de Ecología y Biodiversidad (IEB), Santiago, Chile. ³Programa Austral Patagonia, Instituto de Ciencias Marinas y Limnológicas, Universidad Austral de Chile, Campus Isla Teja, Valdivia, Chile. ⁴Ecole des Hautes Etudes en Sciences Sociales/Centre National de la Recherche Scientifique Institut Interdisciplinaire d'Anthropologie du Contemporain, Paris, France.

*Corresponding author. Email: pmarquet@bio.puc.cl

REFERENCES AND NOTES

1. D. Collins, "Activists outraged that sacred Chilean island is listed for sale for \$20m," *The Guardian* (2020).
2. R. Huckle-Gaete, L. P. Osman, C. A. Moreno, K. P. Findlay, D. K. Ljungblad, *Proc. R. Soc. London Ser. B* **271**, S170 (2004).
3. R. Reyes-Arriagada, P. Campos-Ellwanger, R. P. Schlatter, C. Baduini, *Biodivers. Conserv.* **16**, 913 (2007).
4. R. Álvarez, M. Navarro, in *Conservando el Mar de Chiloé, Palena, y Guaitecas*, R. Huckle-Gaete, P. Lo Moro, J. Ruiz, Eds. (Universidad Austral de Chile, 2010), pp. 65–123 [in Spanish].
5. Climate Action Tracker, Chile, Pledges and Targets (2020); <https://climateactiontracker.org/countries/chile/pledges-and-targets/>.
6. P. A. Marquet et al., Eds., "Biodiversidad y cambio climático en Chile: Evidencia científica para la toma de decisiones" (Comité Científico COP25; Ministerio de Ciencia, Tecnología, Conocimiento e Innovación, 2019) [in Spanish].
7. J. León-Muñoz, M. A. Urbina, R. Garreaud, J. L. Iriarte, *Sci. Rep.* **8**, 1330 (2018).
8. V. Häussermann et al., *PeerJ* **5**, e3123 (2017).
9. M. Rodell et al., *Nature* **557**, 651 (2018).
10. S. Lutz, R. Barnes, T. Kurvitis, "Fish carbon: Exploring marine vertebrate carbon services" (GRID-Arendal, Arendal, Norway, 2014), p. 36.
11. L. Hiriart-Bertrand, J. Silva, S. Gelcich, *Ocean Coast. Manag.* **193**, 105233 (2020).
12. F. Araos et al., *Coast. Manag.* **48**, 289 (2020).

10.1126/science.abf1962

Fisheries rely on threatened salt marshes

Salt marsh ecosystems and the seascapes in which they are embedded serve as critical habitats for species harvested by fisheries (1), which provide food and

economic security for hundreds of millions of people (2). Historical marsh losses coupled with increasing pressures from coastal development and climate change place these intertidal ecosystems and surrounding uplands under growing threat (3). Preventing further losses of salt marshes and associated fisheries production will require greater public awareness and difficult choices in coastal policy and management, underpinned by greater understanding of marsh function.

Quantifying the value of salt marsh habitat to fisheries production is challenging. Many fisheries species feed and shelter in the salt marsh only as juveniles, and it is difficult to assess the marsh's effect once they have moved to a new location (1). It is also unclear how marsh landscape fragmentation under sea level rise will affect fisheries; it may boost fishery production, at least temporarily (4), but it could also disrupt food web processes that support fisheries (5).

Projections of marsh expansion offer hope (6) but are largely dependent on changes in coastal watershed management. For instance, human development may prevent marshes from migrating upland with sea level rise and thus lead to marsh drowning (7). Adequate sediment supply is also essential for marsh resilience, but many coastal areas in the world are sediment-starved (8). Much effort has been made to restore natural riverine flow and other sources of sediment delivery into marshes, although such efforts may have negative impacts on the very fisheries these marshes support (9).

To design effective policies for salt marsh restoration and conservation that protect fisheries production, we need to better understand the role of salt marshes. Researchers should continue to explore the fundamental linkages between salt marshes and fisheries (10), the marsh habitat value within the context of the interconnected and increasingly urbanized mosaic of coastal ecosystems, and the value of salt marshes created by upland transgression and active engineering. Restoration and conservation planning must take a long-term view that specifically recognizes sea level rise and its interaction with other anthropogenic stressors.

Ronald Baker^{1*}, Matthew D. Taylor², Kenneth W. Able³, Michael W. Beck⁴, Just Cebrían⁵, Denise D. Colombano⁶, Rod M. Connolly⁷, Carolyn Currin⁸, Linda A. Deegan⁹, Ilka C. Feller¹⁰, Ben L. Gilby¹¹, Matthew E. Kimball¹², Thomas J. Minello¹³, Lawrence P. Rozas¹⁴, Charles Simenstad¹⁵, R. Eugene Turner¹⁶, Nathan J. Waltham¹⁷, Michael P. Weinstein¹⁸, Shelby L. Ziegler¹⁹, Philine S.E. zu Ermgassen²⁰, Caitlin Alcott²¹, Scott B. Alford²², Myriam A. Barbeau²³, Sarah C. Crosby²⁴, Kate Dodds²⁵, Alyssa Frank¹, Janelle Goeke²⁶, Lucy

NEXTGEN VOICES: SUBMIT NOW

Defining events of 2020

Add your voice to *Science*! Our new NextGen Voices survey is now open:

What new word or phrase would you add to the dictionary to help scientists explain the events of 2020 to future generations?

To submit, go to www.sciencemag.org/nextgen-voices

Deadline for submissions is 20 November. A selection of the best responses will be published in the 1 January 2021 issue of *Science*. Anonymous submissions will not be considered.

A. Goodridge Gaines¹¹, Felicity E. Hardcastle¹¹, Christopher J. Henderson¹¹, W. Ryan James²⁷, Matthew D. Kenworthy²⁸, Justin Lesser²⁷, Debbrota Mallick¹, Charles W. Martin²², Ashley E. McDonald²⁶, Catherine McLuckie²⁹, Blair H. Morrison¹, James A. Nelson²⁷, Gregory S. Norris²³, Jeff Ollerhead³⁰, James W. Pahl³¹, Sarah Ramsden¹, Jennifer S. Rehage³², James F. Reinhardt³³, Ryan J. Rezek³⁴, L. Mark Risse³⁵, Joseph A.M. Smith³⁶, Eric L. Sparks^{37,38}, Lorie W. Staver³⁹

¹Department of Marine Sciences, University of South Alabama, Dauphin Island Sea Lab, Dauphin Island, AL 36528 USA. ²Port Stephens Fisheries Institute, New South Wales Department of Primary Industries, Nelson Bay, NSW 2315, Australia. ³Department of Marine and Coastal Sciences, Rutgers, State University of New Jersey, New Brunswick, NJ 08901, USA. ⁴Institute of Marine Sciences, University of California, Santa Cruz, CA 95062, USA. ⁵Northern Gulf Institute, Mississippi State University, Stennis Space Center, MS 39529, USA. ⁶Department of Environmental Science, Policy, and Management, University of California, Berkeley, Berkeley, CA 94720, USA. ⁷Australian Rivers Institute—Coast & Estuaries, School of Environment and Science, Griffith University, Gold Coast, QLD 4222, Australia. ⁸National Oceanic and Atmospheric Administration (NOAA) National Centers for Coastal Ocean Science, Beaufort, NC 28516, USA. ⁹Woodwell Climate Research Center, Falmouth, MA 02540, USA. ¹⁰Smithsonian Environmental Research Center, Edgewater, MD 21037, USA. ¹¹School of Science and Engineering, University of the Sunshine Coast, Maroochydore, QLD 4558, Australia. ¹²Baruch

Marine Field Laboratory, University of South Carolina, Georgetown, SC 29442, USA. ¹³NOAA Fisheries, Southeast Fisheries Science Center, Galveston, TX 77551, USA. ¹⁴NOAA Fisheries, Estuarine Habitats and Coastal Fisheries Center, Lafayette, LA 70506, USA. ¹⁵School of Aquatic and Fishery Sciences, University of Washington, Seattle, WA 98195–5020, USA. ¹⁶Department of Oceanography and Coastal Sciences, Louisiana State University, Baton Rouge, LA 70803, USA. ¹⁷Centre for Tropical Water and Aquatic Ecosystem Research, and Marine Data Technologies Hub, College of Science and Engineering, James Cook University, Townsville, QLD, 4811, Australia. ¹⁸New Jersey Marine Sciences Consortium, Fort Hancock, Sandy Hook, NJ 07043, USA. ¹⁹Moss Landing Marine Laboratories, Moss Landing, CA 95039, USA. ²⁰Changing Oceans Group, School of Geosciences, Grant Institute, University of Edinburgh, EH9 3FE, UK. ²¹Interfluvie, Hood River, OR 97031 USA. ²²University of Florida, Institute of Food and Agricultural Sciences, Nature Coast Biological Station, University of Florida, Cedar Key, FL 32625 USA. ²³Department of Biology, University of New Brunswick, Fredericton, NB E3B 5A3, Canada. ²⁴Harbor Watch, Earthplace, Inc., Westport, CT 06880, USA. ²⁵Department of Biological Sciences, Macquarie University, Sydney, NSW 2109, Australia. ²⁶Department of Marine Biology, Texas A&M University at Galveston, Galveston, TX 77554, USA. ²⁷Department of Biology, University of Louisiana at Lafayette, Lafayette, LA 70504, USA. ²⁸Department of Marine and Environmental Sciences, Savannah State University, Savannah, GA 31404, USA. ²⁹Department of Environmental Science and Management, University of Newcastle, Ourimbah, NSW 2258, Australia. ³⁰Geography and Environment Department, Mount

Allison University, Sackville, NB E4L 1E4, Canada. ³¹Louisiana Coastal Protection and Restoration Authority, Baton Rouge, LA 70802, USA. ³²Institute of Environment, Florida International University, Miami, FL 33199, USA. ³³NOAA Fisheries, Restoration Center, Silver Spring, MD 20910, USA. ³⁴Department of Earth and Environment, Institute of Environment, Florida International University, Miami, FL 33199, USA. ³⁵University of Georgia Marine Extension and Georgia Sea Grant, Athens, GA 30602, USA. ³⁶U.S. Fish and Wildlife Service, Galloway, NJ, 08205, USA. ³⁷Coastal Research and Extension Center, Mississippi State University, Biloxi, MS 39532, USA. ³⁸Mississippi Alabama Sea Grant Consortium, Ocean Springs, MS 39564, USA. ³⁹University of Maryland Center for Environmental Science, Horn Point Laboratory, Cambridge, MD 21617, USA.
*Corresponding author. Email: rbaker@disl.org

REFERENCES AND NOTES

1. M. W. Beck *et al.*, *Bioscience* **51**, 633 (2001).
2. J. C. Rice, S. M. Garcia, *ICES J. Mar. Sci.* **68**, 1343 (2011).
3. K. B. Gedan, B. R. Silliman, M. D. Bertness, *Annu. Rev. Mar. Sci.* **1**, 117 (2009).
4. E. J. Chesney, D. M. Baltz, R. G. Thomas, *Ecol. Appl.* **10**, 350 (2000).
5. G. A. Hyndes *et al.*, *Biol. Rev.* **89**, 232 (2014).
6. M. Schuercher *et al.*, *Nature* **561**, 231 (2018).
7. J. Fitzsimons, M. W. Beck, L. Hale, K. Leo, C. Gillies, *Ocean Coast. Manag.* **175**, 180 (2019).
8. M. L. Kirwan *et al.*, *Geophys. Res. Lett.* **37**, L23401 (2010).
9. T. J. Mozdzer, E. B. Watson, W. H. Orem, C. Swarzenski, R. E. Turner, *Sci. Tot. Environ.* **743**, 140420 (2020).
10. J. S. Lefcheck *et al.*, *Conserv. Lett.* **12**, e12645, (2019).

10.1126/science.abe9332

PUT YOUR RESEARCH OUT IN FRONT

Submit your research: [cts.ScienceMag.org](https://cts.science.org)

ScienceSignaling
AAAS

ScienceSignaling.org

Twitter: @SciSignal
Facebook: @ScienceSignaling

SAMPLING THE EARLY SOLAR SYSTEM

SCIENCE RESEARCH ARTICLES

- Variations in color and reflectance on the surface of asteroid (101955) Bennu p. 674
 Widespread carbon-bearing materials on near-Earth asteroid (101955) Bennu p. 675
 Bright carbonate veins on asteroid (101955) Bennu: Implications for aqueous alteration history p. 676

SCIENCE ADVANCES RESEARCH ARTICLES

- Heterogeneous mass distribution of the rubble-pile asteroid (101955) Bennu 10.1126/sciadv.abc3350
 Hemispherical differences in the shape and topography of asteroid (101955) Bennu 10.1126/sciadv.abd3649
 Asteroid (101955) Bennu's weak boulders and thermally anomalous equator 10.1126/sciadv.abc3699

By Keith T. Smith¹ and Kip V. Hodges²

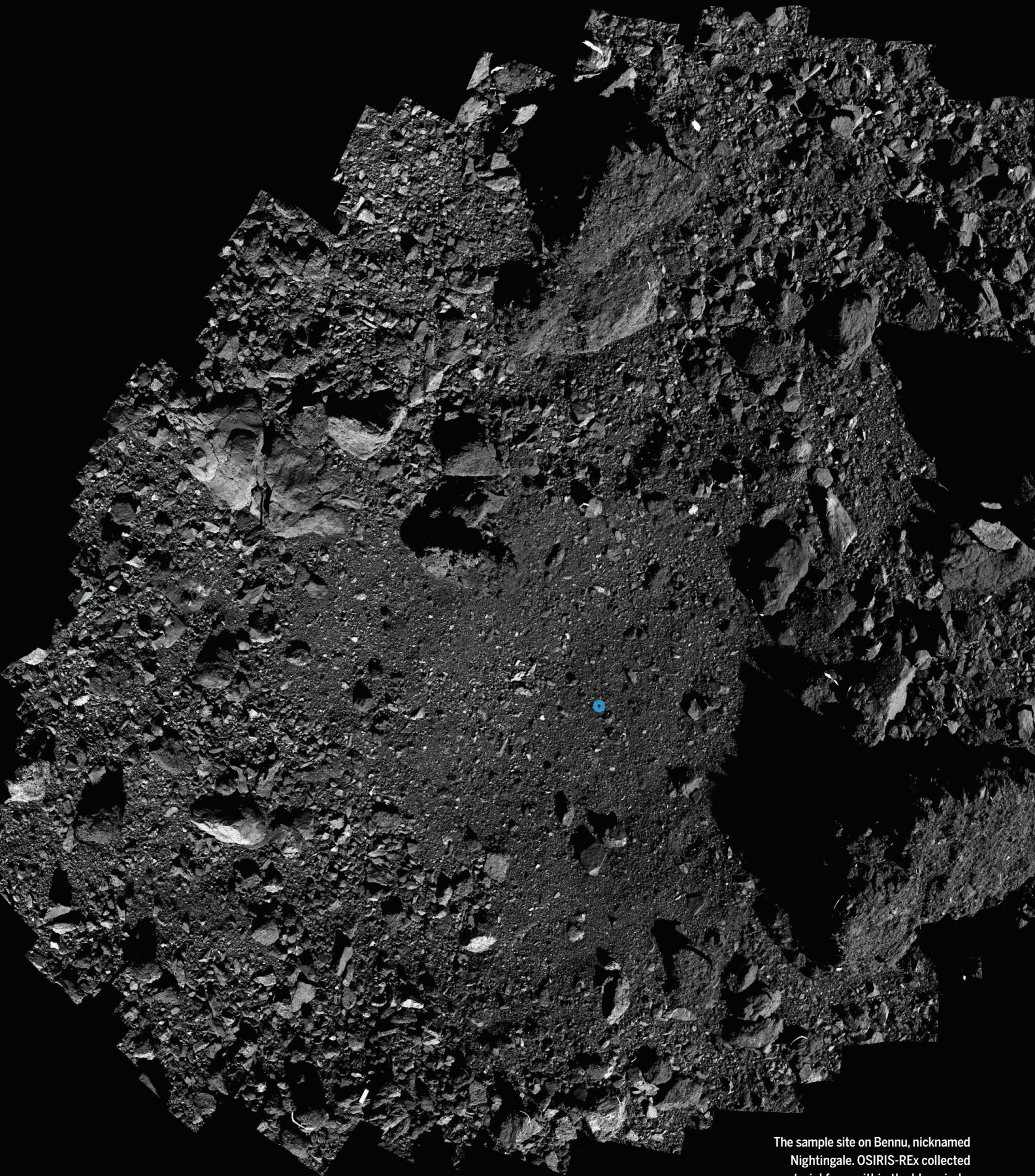
Our knowledge of the Solar System has been greatly advanced by exploration with robotic spacecraft, but there are many limitations on the instruments those can carry. More detailed information can be extracted by analyzing samples of Solar System bodies in Earth laboratories. Meteorites provide some natural samples, but the body they came from is often unknown, and they may not represent typical material. Recent years have seen increased interest in directly retrieving samples from other worlds and bringing them back to Earth, a process known as sample return.

The most primitive carbonaceous asteroids preserve information about the formation and early evolution of the Solar System and thus have been a high priority for sample return. The Hayabusa mission collected 1500 dust particles from asteroid (25143) Itokawa and returned them to Earth in 2010. A successor mission, Hayabusa2, visited asteroid (162173) Ryugu with the goal of obtaining a larger mass of sample, which is due to arrive on

Earth in December 2020. Meanwhile, on 31 December 2018, the OSIRIS-REx (Origins, Spectral Interpretation, Resource Identification, and Security-Regolith Explorer) spacecraft arrived at another primitive asteroid, (101955) Bennu, with plans to return at least 60 grams of material.

Writing in *Science* and *Science Advances*, members of the OSIRIS-REx team present results from the survey and reconnaissance phases of the mission. Detailed maps of surface properties were used to identify promising sites for sample collection and provide the necessary scientific context. These data provide information on Bennu's composition, constrain its formation process, and show how its surface evolved. They show that Bennu's surface material has been modified by exposure to space weathering, contains abundant organic material, and has evidence of past alteration by liquid water. OSIRIS-REx successfully collected its sample of Bennu on 20 October 2020; it is due to arrive on Earth in 2023.

¹Senior Editor, *Science*. ²Deputy Editor, *Science Advances*.



The sample site on Benu, nicknamed Nightingale. OSIRIS-REx collected material from within the blue circle.

RESEARCH ARTICLE SUMMARY

ASTEROIDS

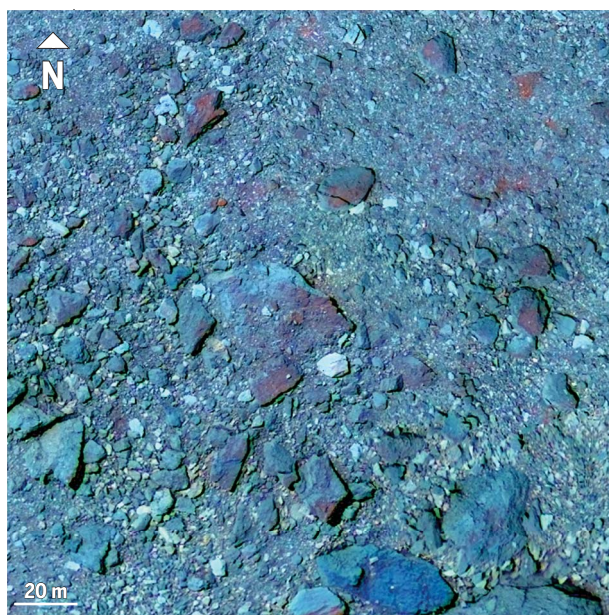
Variations in color and reflectance on the surface of asteroid (101955) Benu

D. N. DellaGiustina*, K. N. Burke, K. J. Walsh, P. H. Smith, D. R. Golish, E. B. Bierhaus, R.-L. Ballouz, T. L. Becker, H. Campins, E. Tatsumi, K. Yumoto, S. Sugita, J. D. Prasanna Deshapriya, E. A. Cloutis, B. E. Clark, A. R. Hendrix, A. Sen, M. M. Al Asad, M. G. Daly, D. M. Applin, C. Avdellidou, M. A. Barucci, K. J. Becker, C. A. Bennett, W. F. Bottke, J. I. Brodbeck, H. C. Connolly Jr., M. Delbo, J. de Leon, C. Y. Drouet d'Aubigny, K. L. Edmundson, S. Fornasier, V. E. Hamilton, P. H. Hasselmann, C. W. Hergenrother, E. S. Howell, E. R. Jawin, H. H. Kaplan, L. Le Corre, L. F. Lim, J. Y. Li, P. Michel, J. L. Molaro, M. C. Nolan, J. Nola, M. Pajola, A. Parkinson, M. Popescu, N. A. Porter, B. Rizk, J. L. Rizos, A. J. Ryan, B. Rozitis, N. K. Shultz, A. A. Simon, D. Trang, R. B. Van Aiken, C. W. V. Wolner, D. S. Lauretta

INTRODUCTION: The color and reflectance of asteroids can be used to infer their compositions and histories. Variations in these spectrophotometric properties are driven by differences in lithology and/or exposure to processes collectively known as space weathering (bombardment by meteoroids and solar wind ions). On anhydrous bodies, such as the Moon and S-type asteroids, space weathering darkens and reddens spectral slopes (where “redder” indicates a more positive slope relative to the solar spectrum) in the visible wavelengths. However, on primitive C-complex asteroids—bodies that may have delivered water and organics to early Earth—the spectral changes that result from space weathering are not well understood. Evidence from meteorites thought to be analogous to C-complex asteroids suggests that either reddening or bluing is possible. Deciphering such changes is necessary to understand the origin and relative exposure age of surface units on primitive Solar System objects.

RATIONALE: The MapCam imager on the OSIRIS-REx (Origins, Spectral Interpretation, Resource Identification, and Security–Regolith Explorer) spacecraft acquired global multispectral images of the C-complex asteroid (101955) Benu in four bands: b' (0.44 to 0.50 μm), v (0.52 to 0.58 μm), w (0.67 to 0.73 μm), and x (0.82 to 0.89 μm). Benu is a rubble pile, made up of the reaccumulated fragments of a larger C-complex progenitor that was blown apart by a catastrophic impact.

Using band ratios and principal components analysis, we mapped Benu's color and reflectance



False-color image mosaic of rubble on asteroid Benu, as observed with the MapCam multispectral imager. Color and reflectance vary between and within boulders, resulting from different exposure ages and innate compositions. Red is the x/v band ratio (indicating redder spectral slopes), green represents relative change in the w band (an indication of composition), and blue is the b'/v band ratio (indicating bluer spectral slopes in the near-ultraviolet region).

tance at a pixel scale of ~ 25 cm. In combination with higher-resolution (~ 5 cm per pixel) panchromatic PolyCam images, we assessed relationships between MapCam spectra and morphologic features on Benu's surface, aiming to understand sources of variation from the average and determine the relative timing associated with color differences.

RESULTS: The surface of Benu has unexpectedly heterogeneous colors distributed on a moderately blue (gently negatively sloped) global surface. Boulders are the dominant source of

heterogeneity and fall into distinct populations on the basis of reflectance. Dark boulders (reflectance of 0.034 to 0.049, encompassing Benu's average reflectance of 0.044) tend to be rougher and rounder, whereas bright boulders (0.049 to 0.074) are smoother and more angular. Variation in color within individual boulders is also apparent; for example, boulder faces that appear to be more recently exposed owing to fracturing are bluer than putatively older faces. Conversely, small reddish craters are observed to overlie blue craters, indicating that the former are younger (more recently exposed material). Benu's smallest craters have a size distribution that indicates that they are also the youngest, and they are redder than the average surface. Many of the larger (older) craters have colors indistinguishable from Benu's average. Crater spectral slopes indicate that terrains with intermediate ages have the bluest near-ultraviolet spectral slopes.

CONCLUSION: The differences in reflectance and texture among boulders indicate that Benu may have inherited distinct lithologies formed at different depths in its larger progenitor asteroid, as well as debris from impactors. The color variations within boulders and among craters suggest that space weathering on Benu does not drive a unidirectional progression from red to blue (or vice versa). Rather, freshly exposed redder surfaces, as exemplified by the small reddish craters, initially brighten in the near-ultraviolet region (i.e., become blue at shorter wavelengths), as exemplified by blue crater rims and fractured boulder faces. Brightening in the visible to near-infrared wavelengths follows, ultimately leading to more moderately blue spectral slopes, consistent with Benu's average. The time scale associated with space weathering-induced color changes ($\sim 10^5$ years) is compatible with previous findings only if Benu's small reddish craters formed under

conditions in which gravity, rather than the strength of the impacted surface, is the dominant influence. This finding offers an indication of cratering physics on small rubble-pile asteroids. ■

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

*Corresponding author. Email: danidg@lpl.arizona.edu
Cite this article as D. N. DellaGiustina *et al.*, *Science* **370**, eabc3660 (2020). DOI: 10.1126/science.abc3660

S READ THE FULL ARTICLE AT
<https://doi.org/10.1126/science.abc3660>

RESEARCH ARTICLE

ASTEROIDS

Variations in color and reflectance on the surface of asteroid (101955) Bennu

D. N. DellaGiustina^{1,2*}, K. N. Burke¹, K. J. Walsh³, P. H. Smith¹, D. R. Golish¹, E. B. Bierhaus⁴, R.-L. Ballouz¹, T. L. Becker¹, H. Campins⁵, E. Tatsumi^{6,7}, K. Yumoto⁷, S. Sugita⁷, J. D. Prasanna Deshapriya⁸, E. A. Cloutis⁹, B. E. Clark¹⁰, A. R. Hendrix¹¹, A. Sen¹⁰, M. M. Al Asad¹², M. G. Daly¹³, D. M. Applin⁹, C. Avdellidou¹⁴, M. A. Barucci⁸, K. J. Becker¹, C. A. Bennett¹, W. F. Bottke³, J. I. Brodbeck¹, H. C. Connolly Jr.¹⁵, M. Delbo¹⁴, J. de Leon⁶, C. Y. Drouet d'Aubigny¹, K. L. Edmundson¹, S. Fornasier^{8,16}, V. E. Hamilton³, P. H. Hasselmann⁸, C. W. Hergenrother¹, E. S. Howell¹, E. R. Jawin¹⁷, H. H. Kaplan¹⁸, L. Le Corre¹¹, L. F. Lim¹⁷, J. Y. Li¹¹, P. Michel¹⁴, J. L. Molaro¹¹, M. C. Nolan¹, J. Nollau⁴, M. Pajola¹⁹, A. Parkinson⁹, M. Popescu^{20,6}, N. A. Porter¹, B. Rizk¹, J. L. Rizos⁶, A. J. Ryan¹, B. Rozitis²¹, N. K. Shultz¹, A. A. Simon¹⁸, D. Trang²², R. B. Van Auken¹, C. W. V. Wolner¹, D. S. Lauretta¹

Visible-wavelength color and reflectance provide information about the geologic history of planetary surfaces. Here we present multispectral images (0.44 to 0.89 micrometers) of near-Earth asteroid (101955) Bennu. The surface has variable colors overlain on a moderately blue global terrain. Two primary boulder types are distinguishable by their reflectance and texture. Space weathering of Bennu surface materials does not simply progress from red to blue (or vice versa). Instead, freshly exposed, redder surfaces initially brighten in the near-ultraviolet region (i.e., become bluer at shorter wavelengths), then brighten in the visible to near-infrared region, leading to Bennu's moderately blue average color. Craters indicate that the time scale of these color changes is $\sim 10^5$ years. We attribute the reflectance and color variation to a combination of primordial heterogeneity and varying exposure ages.

The near-Earth asteroid (101955) Bennu is the target of the OSIRIS-REx (Origins, Spectral Interpretation, Resource Identification, and Security–Regolith Explorer) sample-return spacecraft (1). Before launch,

telescopic observations of Bennu had identified it as a low-albedo object—potentially indicating a carbon-rich composition—with a featureless, gently blue spectrum (blue signifies a negative spectral slope with respect to the solar spectrum, whereas red signifies a positive slope). This resulted in its classification as a blue (B-type) asteroid (2), a subclass of the broader carbonaceous (C-complex) group of small bodies. Bennu is a rubble-pile asteroid (3–6), accumulated from fragments of a larger parent body that was shattered by a catastrophic impact in the inner main asteroid belt ~ 1 billion years ago (3). Bennu eventually migrated from the main belt into its current orbit in near-Earth space (3).

OSIRIS-REx measurements acquired during the initial phases of the mission showed that Bennu is dominated by hydrated clay-bearing minerals (phyllosilicates) and magnetite (7), indicating that water was present on and altered the composition of Bennu's parent body (aqueous alteration). Organic compounds and carbonates have been discovered across the asteroid's surface, supporting the hypothesis that B-type asteroids are carbon rich (8, 9). Carbon-bearing species are optically opaque (as is magnetite) and thus could be responsible for Bennu's low global normal albedo of 0.044 (4, 10). These findings suggest that Bennu's composition may be representative of the primitive bodies that delivered water and organic molecules to early Earth (11).

Although the initial composition of an asteroid influences its global colors, physical properties such as particle size, surface roughness, and porosity can also influence these spectral characteristics, as can duration of exposure to the space environment. The surface colors of airless bodies are expected to be heavily altered by space weathering processes (12, 13), including bombardment by solar wind particles and meteoroids. On anhydrous planetary surfaces, such as the Moon and stony (S-type) asteroids, space weathering darkens and reddens spectral slopes in the visible and near-infrared region (12, 14). However, on primitive carbonaceous asteroids such as Bennu, and their meteorite analogs, space weathering effects are not well understood. Measured colors of primitive asteroids do not show consistent spectral relationships with surface exposure age (13, 15–17), nor do laboratory experiments of simulated space weathering on analogous meteorite and phyllosilicate samples: Some studies indicate that space weathering leads to bluing (13, 18–22), but others find that it leads to reddening (16, 20, 23, 24). This is likely because the initial composition and the physical structure of the materials play a role in the spectral changes observed (16, 20). To determine how space weathering affects low-albedo carbonaceous asteroids, we searched for a correlation between Bennu's colors and the age of its surface features determined from morphology. The spatial distribution and geologic setting of varying colors on Bennu may also aid our understanding of the composition and evolution of the asteroid's surface.

Color observations of Bennu

OSIRIS-REx obtained color observations of Bennu during two hyperbolic flybys on 14 March and 26 September 2019, as part of the Baseball Diamond campaign of the mission's Detailed Survey (25). The OSIRIS-REx Camera Suite (OCAMS) (26) acquired color images using the multispectral MapCam imager, which has four bands in the visible (VIS; 0.40 to 0.70 μm) and near-infrared (NIR; 0.70 to 2.5 μm) wavelengths. The MapCam bands— b' (0.44 to 0.50 μm), v (0.52 to 0.58 μm), w (0.67 to 0.73 μm), and x (0.82 to 0.89 μm)—are similar to those used by telescopic surveys (27) to infer asteroid composition and classify their spectra (28). Most asteroid observations have been limited to unresolved disk-integrated (globally averaged) spectra. We used spatially resolved MapCam color images (pixel scale ~ 25 cm) to investigate reflectance and color across Bennu's surface. To establish relationships between color and surface morphology, we paired these color observations with higher-resolution OCAMS PolyCam panchromatic images (2 to 5 cm per pixel) of the same locations.

We radiometrically calibrated the MapCam and PolyCam images to units of reflectance (also

¹Lunar and Planetary Laboratory, University of Arizona, Tucson, AZ, USA. ²Department of Geosciences, University of Arizona, Tucson, AZ, USA. ³Southwest Research Institute, Boulder, CO, USA. ⁴Lockheed Martin Space, Littleton, CO, USA. ⁵Department of Physics, University of Central Florida, Orlando, FL, USA. ⁶Instituto de Astrofísica de Canarias and Departamento de Astrofísica, Universidad de La Laguna, Tenerife, Spain. ⁷Department of Earth and Planetary Science, University of Tokyo, Tokyo, 113-0033, Japan. ⁸LESIA (Laboratoire d'Etudes Spatiales et d'Instrumentation en Astrophysique), Observatoire de Paris, Université PSL (Paris Sciences & Lettres), CNRS (Centre National de la Recherche Scientifique), Université de Paris, Sorbonne Université, 92195 Meudon, France. ⁹Department of Geography, University of Winnipeg, Winnipeg, MB R3B 2E9, Canada. ¹⁰Department of Physics and Astronomy, Ithaca College, Ithaca, NY, USA. ¹¹Planetary Science Institute, Tucson, AZ, USA. ¹²Department of Earth, Ocean and Atmospheric Sciences, University of British Columbia, Vancouver, BC, Canada. ¹³The Centre for Research in Earth and Space Science, York University, Toronto, ON, Canada. ¹⁴Université Côte d'Azur, Observatoire de la Côte d'Azur, CNRS, Laboratoire Lagrange, Nice, France. ¹⁵Department of Geology, Rowan University, Glassboro, NJ, USA. ¹⁶Institut Universitaire de France (IUF), 1 rue Descartes, 75231 Paris CEDEX 05, France. ¹⁷Smithsonian Institution National Museum of Natural History, Washington, DC, USA. ¹⁸NASA Goddard Space Flight Center, Greenbelt, MD, USA. ¹⁹Istituto Nazionale di Astrofisica (INAF), Osservatorio Astronomico di Padova, Padua, Italy. ²⁰Astronomical Institute of the Romanian Academy, Bucharest, Romania. ²¹The School of Physical Sciences, The Open University, Milton Keynes, UK. ²²University of Hawai'i at Mānoa, Hawai'i Institute of Geophysics and Planetology, Honolulu, HI, USA. *Corresponding author. Email: danidge@lpl.arizona.edu

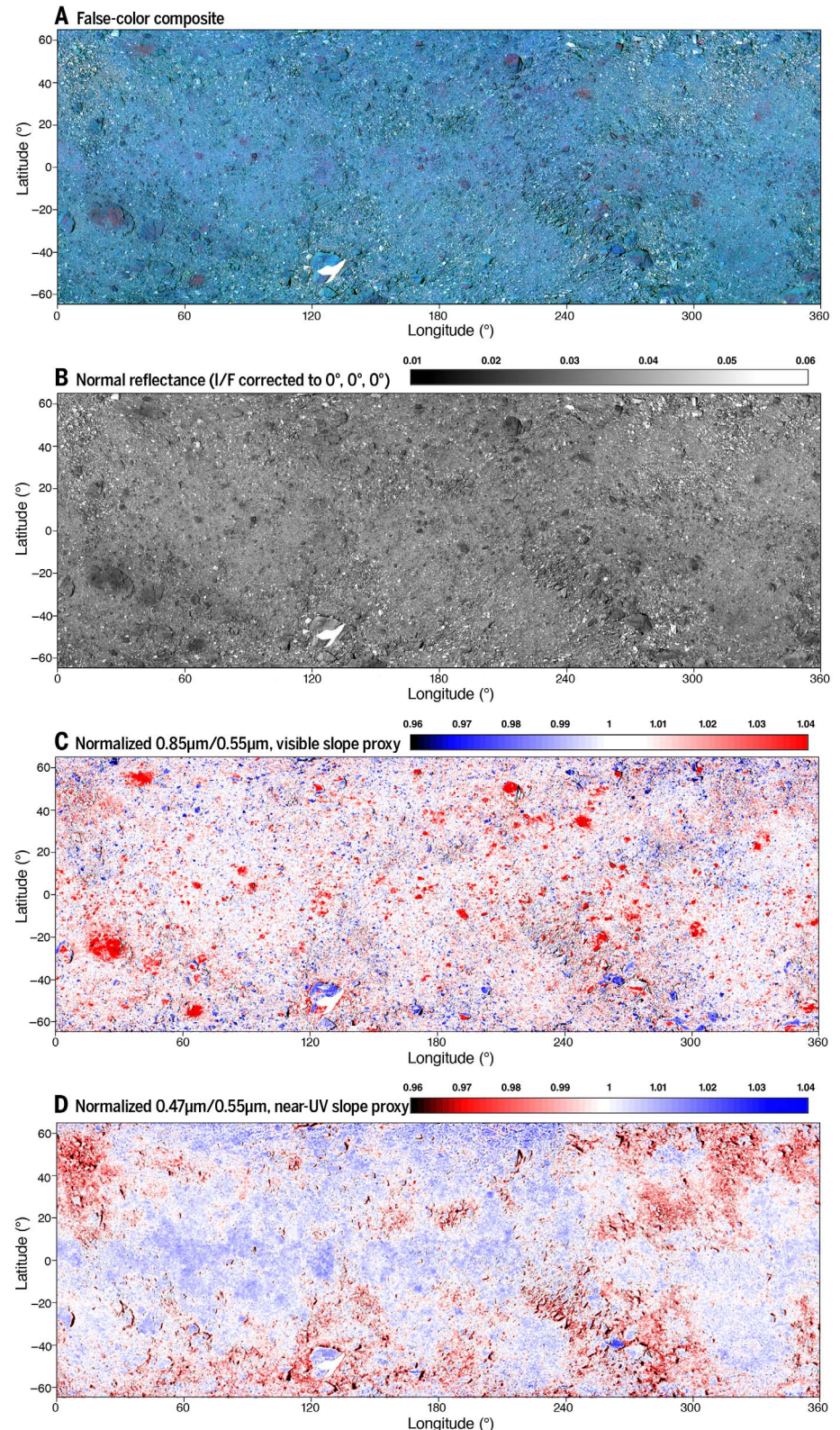
known as radiance factor or I/F (29). MapCam images were corrected to normal viewing conditions (0° solar incidence, 0° solar phase, and 0° observer emission angles) by using a Robotic Lunar Observatory photometric function to

assess albedo differences across the surface (30, 31). We subsequently map-projected and mosaicked the MapCam images using cartographic techniques developed for irregular planetary bodies (23, 31, 32). Band ratios and

principal components analysis (PCA) were used to identify variations in multispectral images and distinguish regions with distinct spectral properties (29). To establish statistically meaningful relationships between color, reflectance,

Fig. 1. Color composite, reflectance, and band ratio maps of Bennu.

(A) False-color red-green-blue (RGB) color model composite overlaid on a $0.55\text{-}\mu\text{m}$ (v band) normal reflectance map of Bennu. Color channels are: red, x/v ($0.85/0.55\text{ }\mu\text{m}$, mid-VIS to NIR spectral slope); green, w band strength (depth at $0.70\text{ }\mu\text{m}$, composition, fig. S5); and blue, b'/v ($0.47/0.55\text{ }\mu\text{m}$, near-UV slope). **(B)** Normal reflectance. **(C)** The x/v band ratio, a proxy for the mid-VIS to NIR spectral slope, where warmer values correspond to redder spectral slopes; values >1 are redder than the global average, and values <1 are less red than the global average. **(D)** The b'/v band ratio, a proxy for the near-UV slope, where higher values correspond to bluer spectral slopes; values >1 are bluer than the global average, and values <1 are less blue than the global average. All maps range from 65°N to 65°S latitude (positive and negative values on the y axes respectively correspond to north and south), 0° to 360°E longitude. The white area near 50°S , 128.5°E represents no data.



and morphological features, we mapped ~1600 boulders and ~700 craters, then extracted their median spectra using established methods (4, 31, 33). We assessed spectral variations that were bounded by irregular morphological

features using centimeter-scale 3D digital terrain models produced with data from the OSIRIS-REx Laser Altimeter (OLA) (34–36). We also compared our findings with near-infrared hyperspectral data acquired at lower spatial

resolution using the OSIRIS-REx Visible and InfraRed Spectrometer (OVIRS) (37).

Bennu's global photometric spectrum is moderately blue (b' to x spectral slope of $-0.1701 \mu\text{m}^{-1}$) in MapCam data (0.44 to $0.89 \mu\text{m}$), but spectral

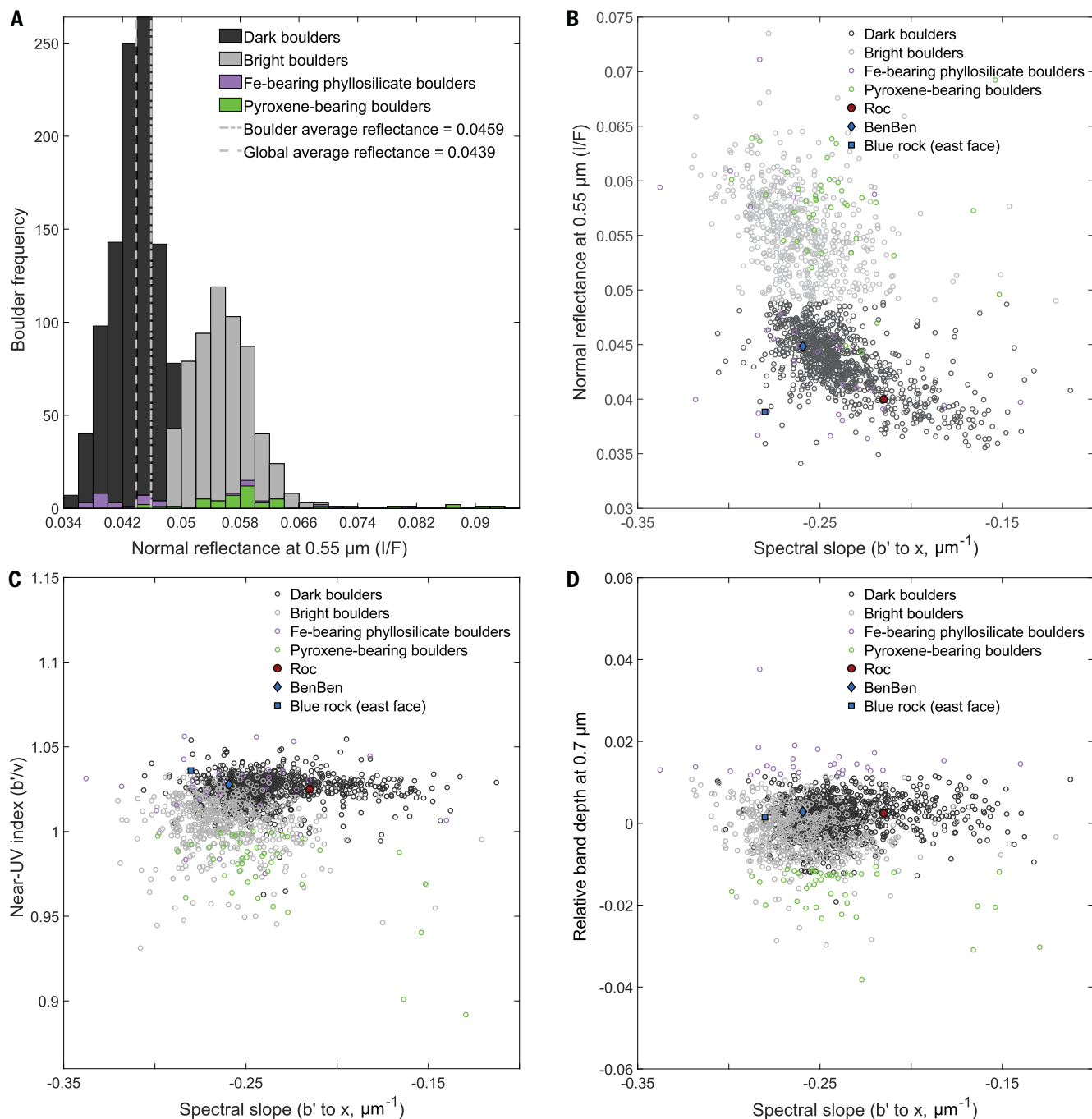


Fig. 2. Variation of boulder color and reflectance on Bennu. (A) Reflectance distribution of boulders ($>5 \text{ m}$) on Bennu. The distribution is multimodal with more than one Gaussian component (31). Shading and colors indicate different classifications of boulders, as indicated in the legend. (B) Normal reflectance versus absolute spectral slope of the same boulder populations; some individual boulders are indicated in the legend. Dark boulders tend to be redder, and their reflectance monotonically decreases with increasing (redder) spectral slopes, whereas bright boulders are bluer and more scattered (31). (C) The near-UV index (b'/v band ratio)

versus the b' to x spectral slope. Dark boulders tend to have a steeper near-UV slope relative to the global average (>1), whereas bright boulders are more often spectrally flat or show a downturn in the near-UV (from the b' to v bands). (D) The relative band depth at $0.7 \mu\text{m}$ (w band) versus the b' to x spectral slope. Some boulders show an absorption feature at $0.7 \mu\text{m}$ (relative w band depth >0), indicative of Fe-bearing phyllosilicates. The boulders named Roc (23.6°S , 25.3°E ; Fig. 3C) and BenBen (46.8°S , 127.5°E) are two of the largest on Bennu; both are dark. The blue rock (39.80°S , 263.02°E) is shown in Fig. 3E.

slopes vary from blue (negative, $-0.25 \mu\text{m}^{-1}$) to red (positive, $0.05 \mu\text{m}^{-1}$) at spatial scales of as small as 2 m (Fig. 1). Bennu's surface exhibits widespread heterogeneity in reflectance (Figs. 1 and 2). PCA shows that the first principal component (PC1) corresponds to albedo, whereas the second principal component (PC2) corresponds to changes in the overall spectral slope (from b' to x), and PC3 indicates variation in the near-ultraviolet (near-UV) region (from b' to v) (figs. S1 to S3).

Bennu's globally blue surface is dominated by a coarse layer of decimeter- to meter-scale rocks with some centimeter-scale particles (regolith), which we refer to as average terrain. The spectral variability is associated with distinct geologic features, including boulders, craters, and areas of mass wasting (rock movement down geopotential slopes), which we used to distinguish color units on the asteroid (Table 1 and fig. S4, A and B). These units provide a framework for classifying commonly observed features on the surface.

Boulders on Bennu

Boulders are the primary source of heterogeneity on Bennu. Some individual boulders have VIS-NIR absorption features at $0.55 \mu\text{m}$ [as previously observed (5)], $0.7 \mu\text{m}$ (Fig. 2D), and near $1 \mu\text{m}$ [as previously observed (38)]. There are also outliers among the boulder population that do not correspond with a specific color unit.

Boulders have a wide range of normal reflectance values [0.032 to at least 0.26 (38)] that are multimodally distributed with two prominent peaks (Fig. 2A). The reflectance of boulders is not normally distributed and is most consistent with four Gaussians (31), or perhaps two or more non-Gaussian components. We refer to boulders as bright or dark based on whether they are brighter or darker than the median reflectance of 0.049, which lies between the two prominent peaks of the reflectance distribution.

Bright boulders have smooth surfaces, typically angular shapes (Fig. 3, A and B), and blue spectral slopes in the mid-VIS to NIR (MapCam v to x bands). However, unlike the average terrain, ~80% of the bright boulders are dark in the near-UV region ($b'/v < 1$; Fig. 2 and fig. S4, A and B). The bright boulders appear to have similar sizes (are well sorted) with diameters <10 m.

Dark boulders (reflectance ≤ 0.049) span a range of visible spectral slopes but are generally redder than the bright boulders in the mid-VIS to NIR wavelengths (Figs. 1 and 2, A and B). The dark boulders are less angular than bright boulders and commonly have rougher, more undulating surface textures (Fig. 3, C and D). They encompass a wide range of sizes (from decimeters to ~95 m) and include all large (≥ 20 m) boulders on the asteroid. Although the average terrain lies between the

Table 1. Adopted color units on Bennu and their distinguishing properties.

Color unit	Distinguishing properties
Bright boulders	Normal reflectance ranges from 0.049 to 0.074. Bluish spectral slopes are slightly steeper than the global average in the mid-VIS to NIR wavelengths (0.52 to $0.96 \mu\text{m}$) but are often spectrally flat or show a downturn in the near-UV wavelengths (0.44 to $0.50 \mu\text{m}$; MapCam b' band). Diameters are <10 m.
Dark boulders	Normal reflectance ranges from 0.034 to 0.049, overlapping with the average reflectance of Bennu (0.044). Slopes tend to be redder in the mid-VIS to NIR wavelengths (0.52 to $0.96 \mu\text{m}$) and often show an upturn in the near-UV wavelengths, consistent with the global average. Diameters range from decimeters to ~95 m.
Fe-bearing phyllosilicate boulders	An absorption feature is present at $0.70 \mu\text{m}$ (determined from a relative band depth $>1\%$ in the MapCam w band). Reflectance spans a wide range, most often overlapping with that of dark boulders.
Pyroxene-bearing boulders	An absorption feature is present beyond $0.89 \mu\text{m}$ (determined from a downturn in the MapCam x band relative to the w band) due to pyroxene (36). Reflectance may reach up to 0.26 (36). These units can occur as discrete boulders or pyroxene-bearing clasts in a dark boulder-like matrix.
Blue units	These boulders and craters show an upturn in the near-UV wavelengths that exceeds that of the global average photometric spectrum, possibly resulting from an absorption at $0.55 \mu\text{m}$.
Reddish craters	These small (<25 -m diameter) craters are $\geq 0.5\sigma$ redder than Bennu's global average (median) and contain material that is not resolved at ~2 cm per pixel.
Breccias	These medium (~5 m) to large (>10 m) boulders have embedded clasts (tens of centimeters) whose spectrophotometric properties are distinct from those of the host matrix, which resembles the dark boulders.
Average terrain	These areas are absent of large boulders (≥ 20 m) and have a photometric spectrum similar to the global average (median).

bright and dark populations, dark boulders and smaller dark particles (presumably formed by boulder breakdown) appear to be the dominant material; this is illustrated by the close correspondence between Bennu's average reflectance (0.0439 ± 0.002) and the reflectance peak of the dark boulder population (0.0450 ± 0.002) (Fig. 2A, gray dashed line).

Color variation is also evident within individual boulders (Figs. 1 and 3). Sometimes this occurs between faces of an individual rock, usually a large dark boulder (Fig. 3, C to H). These examples of intraboulder color variation are associated with apparent exfoliation and fracturing of the rock (39) (Fig. 3, G and

H) and occur in textures akin to weathering rinds—exterior crusts that appear discolored compared with faces that are potentially more recently exposed (Fig. 3, C and G). We also observed boulders that appear to be breccias—that is, objects composed of rock fragments cemented together as a result of large impacts on Bennu's parent body (33, 38). In these boulders, spectrophotometrically distinguishable fragments (clasts) are embedded in a host matrix that has similar texture, reflectance, and color to the dark boulders (Fig. 3, E and F).

A small population of boulders with very high reflectance [up to 0.26 (38)] shows evidence of an absorption feature at $1 \mu\text{m}$ (downturn in

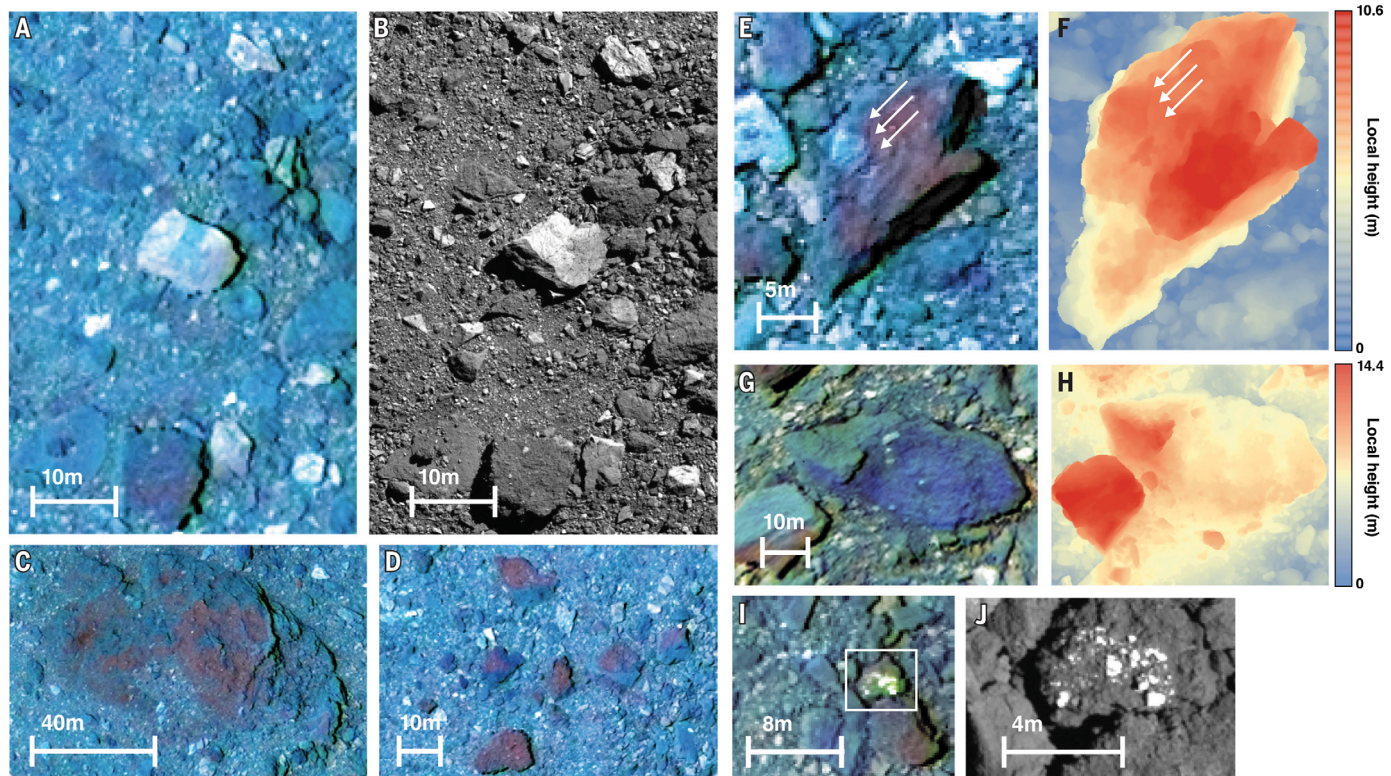


Fig. 3. Examples of boulder color and morphology. All RGB color composite images are shown on the same color scale as in Fig. 1A. (A and B) Bright boulders near 17.76°S, 74.74°E in (A) MapCam (25 cm per pixel) and (B) PolyCam (5.25 cm per pixel) images. The higher-resolution panchromatic PolyCam image (B) shows the angular morphology of bright boulders. (C and D) Dark boulders near (C) 23.6°S, 25.3°E and (D) 3.92°N, 178.98°E, some of which display intraboulder color variation. This includes Roc (C), which is the largest boulder observed on Bennu at 95 m in the longest observable dimension.

(E) A dark boulder with resolvable clasts that appear distinct from the host matrix (4.62°S, 248.95°E). (F) Digital terrain model of the same boulder. The three arrows highlight the same clast in (E) and (F). (G) A rock at 39.80°S, 263.02°E. The darker and bluer portion corresponds to a lower-relief fractured face in the digital terrain model (H). (I) A boulder that appears to contain clasts and be brecciated; the bright clasts with a greenish color signature are indicative of pyroxene in this false-color scale. (J) A higher-resolution PolyCam image (5.25 cm per pixel) of the same boulder. A wider context image is shown in fig. S15.

the x band). OVIRS data indicate that these boulders contain pyroxene (38). Pyroxene-bearing material appears in distinct clasts embedded within larger host rocks whose color and reflectance are similar to those of the dark boulders (Fig. 3, I and J) (38); it also appears in smaller (meter-scale), isolated boulders that do not look brecciated. Pyroxene was probably inherited from Bennu's parent body, where it was implanted by an impactor that may have originated from a fragment of (4) Vesta (38), the differentiated (not primitive) inner-main-belt asteroid visited by the Dawn mission (40). Although their reflectance overlaps with that of the breccias, we separated pyroxene-bearing boulders into their own category (Table 1) based on their distinct spectral shape in the VIS to NIR (v to x ; Fig. 2, B and C).

Some boulders have an absorption feature at 0.7 μ m (absorption depth of 2 to 10%). Similar absorption features have been observed in spectra of primitive asteroids and carbonaceous meteorites, where they were attributed to iron

in some clay-bearing phyllosilicates (41). Boulders with the deepest absorptions at this wavelength span a wide range of reflectance (0.036 to 0.081); however, most (~60%) dark boulders tend to have a shallow (~1%) 0.7- μ m absorption. Although this is at the limit of the 1% relative precision of OCAMS data (29), it is spatially coherent with individual boulders, giving us confidence that the feature is real (fig. S5).

Color and surface processes

Some areas of the surface are brighter in the near-UV region (i.e., show an upturn in the near-UV b' band; $b'/v > 1.01$, or 1% greater than average); we refer to these areas as blue units. One of the bluest contiguous surfaces on Bennu is the eastern side of the boulder shown in Fig. 3G. Digital terrain models show that this rock face is recessed relative to the western face of the boulder (Fig. 3H), and color data indicate that the recessed face has a steeper blue spectral slope. Conversely, the western face of the boulder is brighter and has a more neutral spectral slope, similar to the global average

(fig. S4C). The bluer, recessed eastern face may be a fresher surface, more recently exposed to the space environment by thermal fracturing or exfoliation—ongoing surface processes on Bennu (39, 42). Similar patterns are observed on other boulders that appear to have been recently exfoliated (as indicated by layers with different relief). We also observe bluer-than-average near-UV slopes on the high-elevation rims (10 to 20 m higher than surrounding terrains) of equatorial craters that show recent indications of mass wasting (Fig. 4, A and B) (43).

The correspondence between blue units and potentially less weathered surfaces suggests that enhanced near-UV reflectance is related to younger exposure ages on Bennu. The blue units are correlated with the relative band depth at 0.55 μ m (fig. S6), indicating that the upturn observed in the near-UV index may be the result of an absorption feature at 0.55 μ m. This absorption feature on Bennu has previously been attributed to magnetite (5), which is also detected in thermal emissivity spectra from the OSIRIS-REx Thermal

Emission Spectrometer (OTES) (7). However, this absorption feature could also result from graphitized carbon (44), and both magnetite and graphite are associated with space weathering (discussed further below).

The color of the largest craters (>100 m) on Bennu is indistinguishable from that of the average terrain. However, many small craters (≤ 25 m) are redder than Bennu's average by $\geq 0.5\sigma$ in the near-UV to NIR, where σ is the full-width at half maximum of the global distribution of b' to x spectral slopes (Figs. 4, C and D, and 5, A and B). We refer to these as small reddish craters. By contrast, we identified no craters bluer than Bennu's global average spectral slope by $\leq 0.5\sigma$ (Fig. 5A).

The size-frequency distribution of reddish craters implies that they are one of the youngest components of the global crater population (Fig. 5C). If so, we expect that reddish craters are among the youngest surface features on Bennu. The absolute spectral slopes of the reddish craters appear to correspond with crater size (Fig. 5, A and B), and these craters are also darker than the global average (Fig. 5D). In PolyCam images, the reddish craters display a texture distinct from that of the bulk of Bennu's surface: unresolved at the pixel scale (i.e., <5 cm), indicating fine-particulate material (fig. S7). The largest examples of reddish craters exist at mid- to high latitudes (poleward of $\pm 20^\circ$), including the crater selected

as the primary OSIRIS-REx sample collection site, Nightingale (56°N , 42°E).

Evidence of freshly exposed material on Bennu

Reddish craters and blue units on Bennu both correspond to apparently young exposure ages (Figs. 3, G and H, and 4, C and D). To resolve this apparent contradiction, we considered crater production rates and size-scaling laws, stratigraphic relationships, and potential contributors to spectral slopes other than exposure age.

The size-frequency distribution of the reddish craters has a power-law index of -2.1 ± 0.4 (31). The power-law index of the reddest subset of these craters (b' to x spectral slopes $\geq 1\sigma$ from the global median value, i.e., $\geq -0.0698 \mu\text{m}^{-1}$) is -2.3 ± 0.6 (31). These values are close to that of the expected production of craters in the present-day main asteroid belt and near-Earth space over the past 100,000 years (predicted power-law index between -2.6 and -2.7 ; Fig. 5C) (45–48). The global crater population has a different power-law index of -1.1 ± 0.1 (31). At small diameters, the distribution of the reddish craters does not deviate from that of the expected crater production (Fig. 5C), unlike the global population of craters on Bennu (33) and nearly all other closely studied small bodies (49). A mismatch at small diameters between observed crater populations and the expected crater production has been attributed to erasure processes (49), which efficiently erode and diminish small craters. The correspondence between the expected production of the redder craters on Bennu and their actual distribution supports a young age: Unlike the global crater distribution, they have not yet experienced substantial erasure.

We estimated the time required to produce the craters on Bennu using a crater production rate and a scaling law for crater size from impactor size (31). Assuming that the redder craters have formed since Bennu's arrival in the inner Solar System, we adopted the established crater production rate from the near-Earth object population (47, 48). Crater scaling relationships, however, are less certain owing to unknown material properties and the structure of rubble-pile asteroids. The Small Carry-on Impactor (SCI) experiment performed on asteroid (162173) Ryugu by the Hayabusa2 mission (50) showed that small impactors (~ 30 cm) can produce craters consistent with scaling laws that depend on gravity; craters formed in this gravity regime can be many times larger than the sizes expected from scaling laws governed by target material strength. The sizes of the reddish craters on Bennu are smaller than or similar to the diameter of the SCI crater (~ 18 m) (50). If gravity dominates the cratering process for small craters on Bennu (51), the reddest subset of these craters is less than 10^5 years old

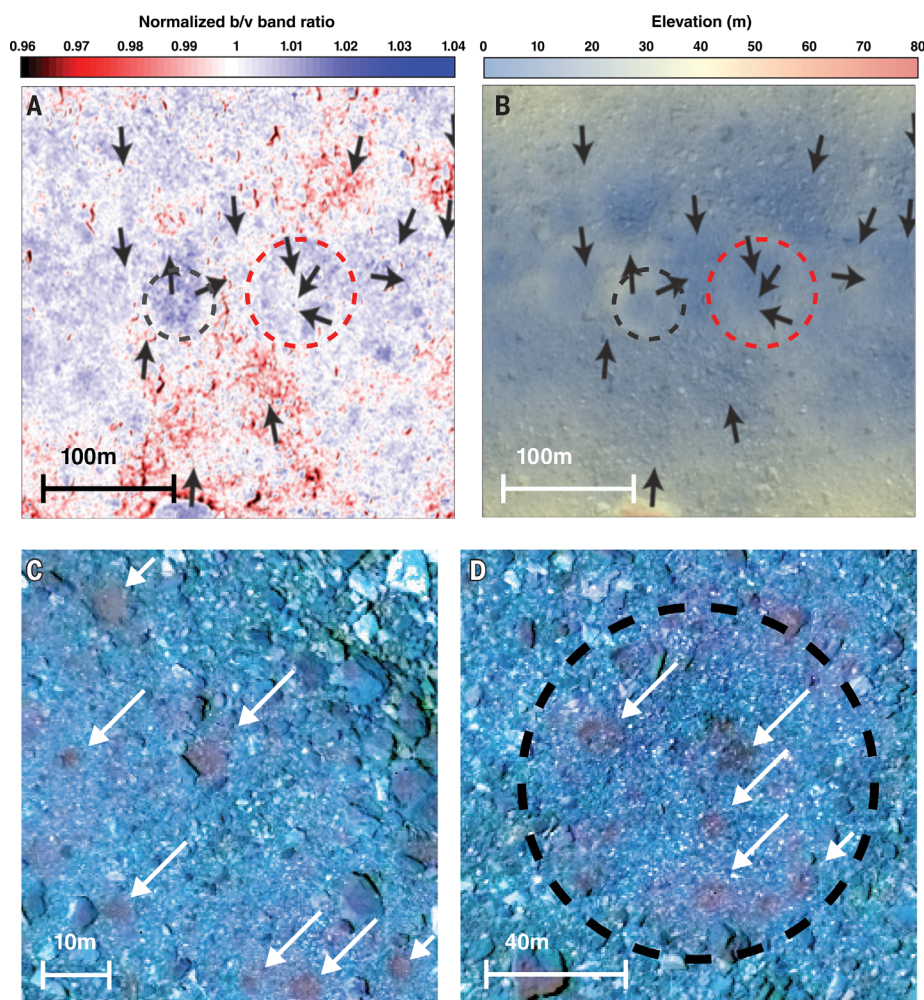


Fig. 4. Examples of crater color. (A) The b'/v band ratio map of equatorial craters at 3.05°S , 128.79°E (black dashed circle) and 1.06°S , 152.75°E (red dashed circle), which have a relatively high near-UV index and are considered blue units. (B) An elevation map of the same equatorial craters in (A). These blue units correspond with areas that show recent indications of mass movement away from regionally high elevations, such as from crater rims. Black arrows correspond to locations of previously mapped mass movement (43). (C and D) RGB color composite images of small craters (indicated by white arrows), shown on the same color scale as in Fig. 1A. As indicated by their color, these craters are consistently redder than Bennu's average terrain, with positive to slightly blue b' to x spectral slopes. The blue crater indicated by the dashed black circle in (A), (B), and (D) has been overprinted by several smaller reddish craters.

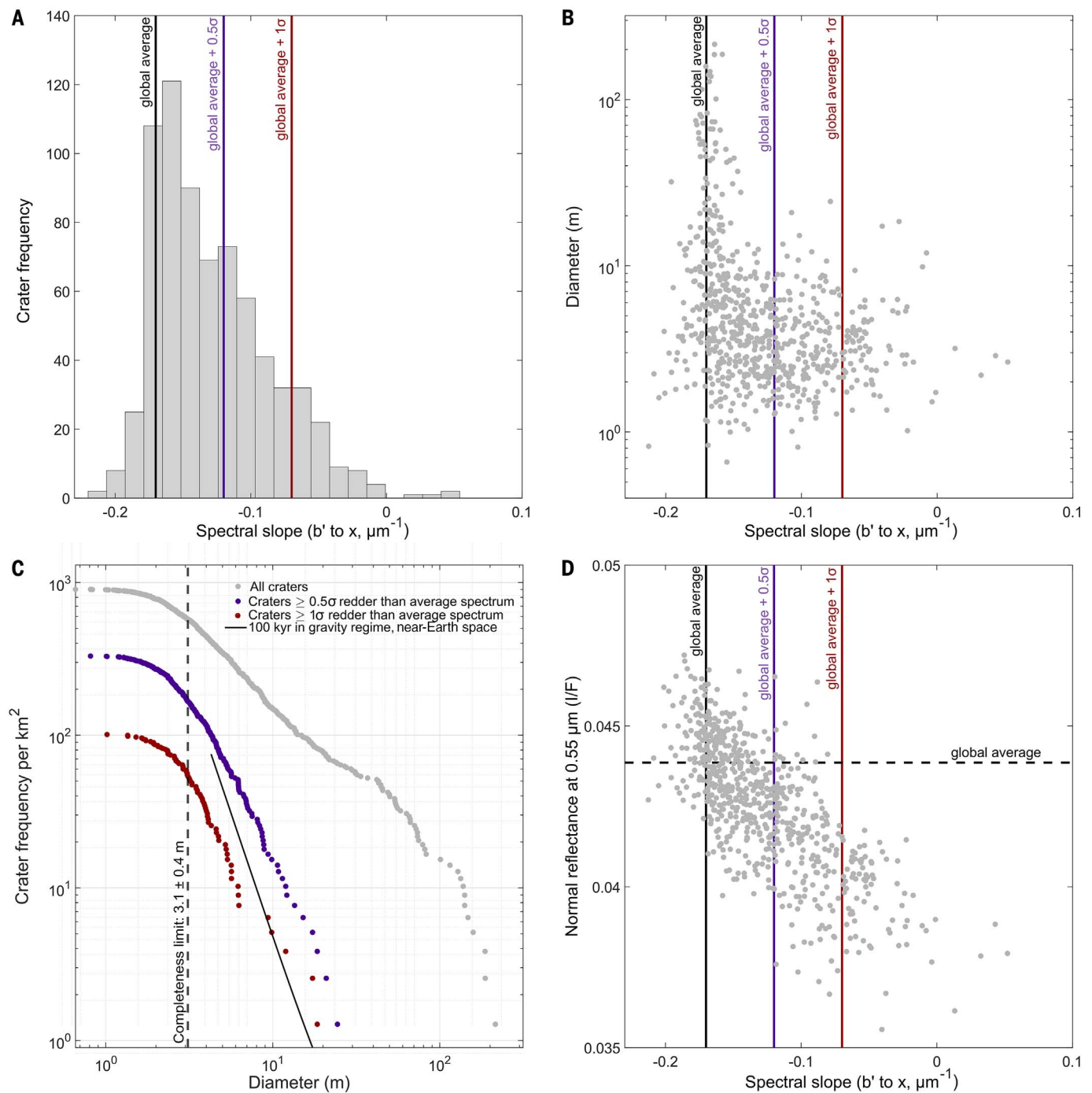


Fig. 5. Color, reflectance, and size distribution of Bennu's craters. (A) The b' to x spectral slope distribution of craters on Bennu, which are not normally distributed. (B) Crater diameter as a function of the crater's median b' to x spectral slope. Craters have more negative spectral slopes at higher size and frequency; this suggests that redder (more positively sloped) craters are younger. (C) The crater size frequency distribution on Bennu for craters classified on the basis of their b' to x spectral slopes. The black line shows the expected crater production for 100,000 years in near-Earth space (cumulative power-law index of -2.7), assuming cratering in the gravity regime (51). The size-frequency distribution of

the reddish craters (purple circles) is more consistent with the black line at small diameters than that of the global crater population (gray circles). The reddest subset of craters (red circles) falls below the black line and appears to have formed more recently than 100,000 years ago. (D) The normal reflectance of craters on Bennu as a function of b' to x spectral slope. Like that of dark boulders, the reflectance of craters monotonically decreases with increasing (redder) spectral slopes (31). Supporting information for the relationship between spectral slope and crater size and frequency is shown in fig. S16.

(Fig. 5C). This is consistent with the expected time scale of space weathering on near-Earth asteroids ($\sim 10^5$ years) based on returned samples of asteroid Itokawa (52) and laboratory experiments on primitive meteorites (20). Thus, it is plausible that the reddish craters are the most recent areas of surface exposure and represent the least weathered material on Bennu.

Although recent surface exposure may explain the reddish craters on Bennu, in carbonaceous meteorites, redder spectral slopes in the VIS-NIR can also arise from fine particle sizes ($<100\ \mu\text{m}$) (53). The smooth and unresolved appearance of the reddish craters on Bennu suggests that they possess finer-scale regolith (fines) than elsewhere on the surface (fig. S7). However, small craters on boulders also appear redder than the global average photometric spectrum, indicating that particle size is not the only factor (fig. S8). When the electrostatic forces acting on particles exceed that of gravity and cohesion—as is often the case for micrometer-sized grains on asteroids—lofting occurs and preferentially removes fines (54). Simulations of electrostatic lofting show that detachment and escape of submillimeter particles is feasible on Bennu and likely occurs shortly after the formation of such fines (54). Although particle sizes $<100\ \mu\text{m}$ could contribute to reddening on Bennu, we anticipate that the loss of fines takes place on shorter time scales than the age of the reddish craters. However, if micrometer-scale particles are closely associated with freshly exposed materials, they could also produce the observed reddening.

Longer infrared wavelengths also indicate younger exposure ages for reddish craters. The shape of the $2.7\text{-}\mu\text{m}$ hydration feature, which is ubiquitous on Bennu (7), is sharper and shifts to shorter wavelengths within reddish craters, including the Nightingale sample site (fig. S9A). This is consistent with laboratory space weathering experiments performed on carbonaceous meteorites (20), which show that the minimum of the hydration feature at $\sim 2.7\ \mu\text{m}$ moves toward longer wavelengths with prolonged exposure to space weathering processes. Thus, sharper band features with shorter minimum wavelengths within craters may signify more recent exposure. The blue equatorial crater at 1.06°N , 152.75°E (Fig. 4, A and B) appears to have a sharper absorption feature but a band minimum similar to the global average (fig. S9B), indicating that it may have an intermediate exposure age. Unlike the spectral slope, the shape of this spectral signature is not expected to be influenced by particle size (20).

The exposure age is less clear for the bluer-than-average, apparently broken boulder face in Fig. 3, G and H. We can set some constraints from the blue units associated with equatorial craters. Since its time in near-Earth space, Bennu's rotation rate has accelerated in response to surface scattering of sunlight and

the emission of its own thermal radiation (10, 55), consequences of the Yarkovsky-O'Keefe-Radzievskii-Paddack (YORP) effect. The locations of two of the bluest craters (1.06°S , 152.75°E and 3.05°S , 128.79°E) correspond to an isolated equatorial region that experienced increased surface accelerations as Bennu's spin period decreased from 5 hours to the present-day 4.3 hours (fig. S10) (10, 55), suggesting that this area has experienced surface mass movement at some point in the past 200,000 years (43). Thus, we infer that the exposure age of blue units is $\sim 200,000$ years.

Blue unit craters have also been overprinted by small reddish craters (Fig. 4D); we therefore presume that they are composed of the same underlying material and that the blue unit craters are older than the reddish craters. This color trend of bluing with exposure time is consistent with some spectral studies of space weathering of carbonaceous meteorites (18–22) but not previous surface-resolved observations of asteroids.

We suggest that nonlinear space weathering occurs on Bennu's surface. In this scenario, freshly exposed material is initially redder than the bulk of Bennu in the near-UV to NIR wavelengths (b' to x) (first stage). During early space weathering, these surfaces brighten more rapidly in the near-UV (b') than in the mid-VIS to NIR (v to x), thereby increasing the b'/v band ratio (middle stage). This near-UV bluing may result from the deepening of an absorption near $0.55\ \mu\text{m}$, implying that magnetite or graphite abundance increases during the early stages of space weathering. Eventually, however, these color differences neutralize as the surface is brightened across the wavelengths observed by MapCam, and the downturn at $0.55\ \mu\text{m}$ (v band) diminishes, leading to the gently blue spectral slope that characterizes Bennu's global photometric spectrum and the oldest craters (final stage).

This space weathering progression is demonstrated by craters that have been categorized on the basis of their spectral slope (Fig. 6). The first and middle stages are illustrated by the small reddish craters overprinted on blue equatorial craters (Fig. 4D). The middle and final stages may also be illustrated by the boulder in Fig. 3G, whose higher-relief, presumably older western face is brighter and has a more neutral spectral slope than the more recently exposed, bluer face. The final stage is illustrated by Bennu's average terrain.

The orientation of intraboulder color variation provides further evidence of space weathering on Bennu. We examined 220 boulders with sizes $\geq 5\ \text{m}$ in the equatorial region (20°S to 20°N), where MapCam images have the most consistent viewing conditions, and calculated the azimuthal angle that maximizes blue-to-red color variation across each boulder (31). Figure S11 shows the latitude of boulders

with large color variation (b' to x slope difference of $>0.05\ \mu\text{m}^{-1}$ between the two regions) as a function of the azimuthal angle of the blue-to-red direction (31). Bluer faces are preferentially oriented toward the equator (fig. S12). This may relate to solar exposure (e.g., heating or irradiation) and/or meteoroid bombardment, all processes that would preferentially affect equatorial latitudes (42, 56). We observed this intracolor variation only on dark boulders, perhaps owing to their larger sizes compared with other boulder types, or a different initial composition.

Mechanisms for space weathering on Bennu

Dark, optically opaque minerals, such as nanophase iron, graphitized carbon, sulfides, and magnetite, are commonly produced by space weathering of primitive materials (19, 23, 44, 57). Laboratory studies have shown that phyllosilicates intimately mixed with certain carbon species (including graphite) and magnetite can lead to bluer and darker spectral slopes in the visible wavelengths (44) (fig. S4, C and D). The link between bluing opaque minerals and space weathering implies that blue spectral slopes are the result of a mature, weathered regolith on primitive B-type asteroids, although they cannot fully explain the brightening in the near-UV to NIR observed on Bennu.

As discussed above, studies of primitive asteroids and meteorites have contradictorily predicted both bluing and reddening with increased space weathering (13–24). Ion and meteoroid bombardment lead to darkening and reddening of anhydrous silicate planetary surfaces; this has previously been attributed to the accumulation of nanophase and larger metallic iron particles, which form in response to the space environment (58). However, contrary to the bluing and darkening predicted from the production of graphitized carbon and magnetite, the formation of nanophase iron in low-albedo asteroids such as Bennu may have a reddening and brightening effect, owing to the higher reflectance of nanophase iron than that of the primitive materials found in carbonaceous meteorites (59).

Space weathering trends observed on Bennu corroborate earlier studies (13, 44) that found that primitive asteroids, though dark relative to the asteroid population, are brighter in the UV than their primitive meteorite counterparts—that is, they are spectrally bluer at shorter wavelengths. Ion bombardment of primitive low-albedo meteorites in the laboratory leads to spectral bluing and brightening (20). These spectral changes are attributed to the process of carbonization, whereby hydrogen is lost and crystalline carbon structures, such as graphite, are formed (57). Irradiation experiments on complex hydrocarbons have shown that carbonization can induce metal-like optical properties in carbonaceous material, which leads

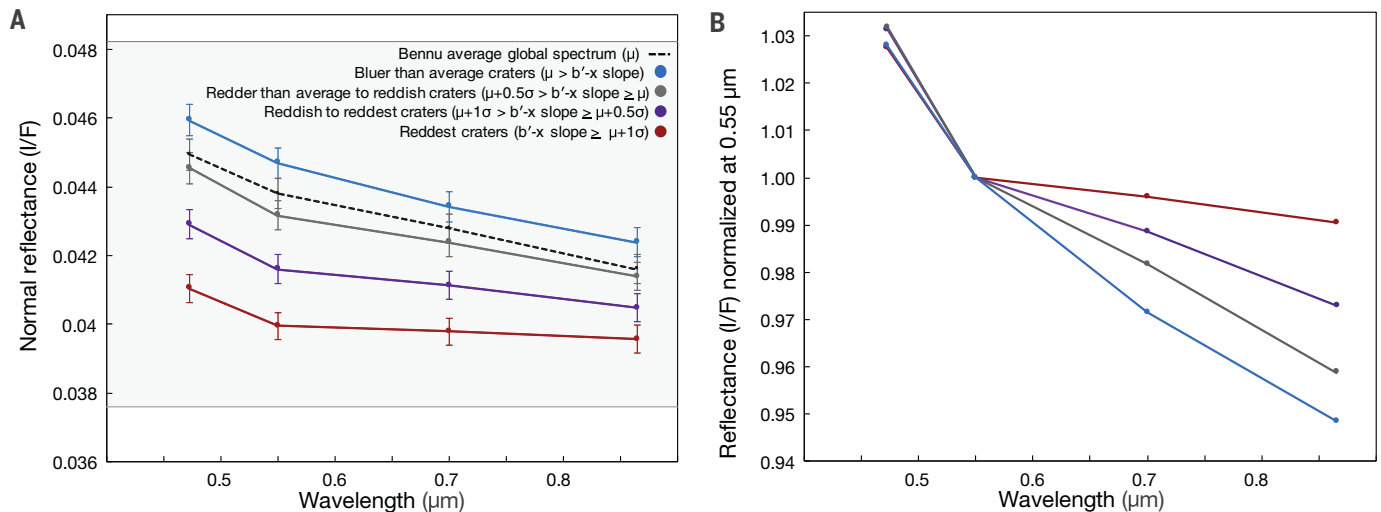


Fig. 6. Evolution of crater colors from MapCam data. Older craters have bluer overall spectra from b' to x (Fig. 5); here we show variations in each filter separately. **(A)** The average absolute reflectance spectra of all craters within a given b' to x slope range (see Fig. 5A). The light gray box encompasses the range of absolute radiometric uncertainty (29), whereas error bars show the relative precision of OCAMS measurements (26). Craters become brighter as their spectral slopes steepen. **(B)** The same reflectance spectra shown in (A),

normalized at 0.55 μm . The progression from the mid-VIS to NIR (v to x bands) dominates the evolution of crater spectra, which mature toward more negative slopes. In the near-UV, spectra with intermediate b' to x slopes steepen in a subtle but statistically significant way (31). This may result from more rapid brightening in the near-UV relative to longer wavelengths, a deepening of absorption feature at 0.55 μm (v band), or both. As a result, crater spectra show a non-unidirectional change near the b' band as they age.

to bluing and brightening (17, 57). We propose that space weathering-induced carbonization or magnetite formation may have influenced the surface colors on Bennu, especially in dark materials that become bluer (from the v to x bands) and brighter (across all bands) with increased exposure age.

Evidence for parent body heterogeneity

Although space weathering may influence the spectral slopes on Bennu, the disparate boulder populations that we identify—which differ in terms of their reflectance, texture, and size—suggest that some of Bennu's heterogeneity was inherited from its parent body. Dark boulders show a monotonic relationship of decreasing reflectance with increasing (redder) spectral slope, which is distinguishable from that of their bright counterparts (Fig. 2B) (31). Because space weathering is likely controlled by initial texture and composition, the difference in spectral slope trends between dark and bright boulders could result from the maturation of geologically distinct materials. Likewise, the less varied spectrophotometric properties of the bright boulders suggest that they are less susceptible to modification from space weathering or change on a different time scale (Fig. 2).

Heterogeneity in boulder reflectance on Bennu may be the result of distinct rock types that fragment differently, leading to an observed difference in their size-frequency distributions (4). Different compositions of the dark versus bright boulders may account for the correlation between reflectance and thermal prop-

erties for boulders on Bennu (60). The dark boulders have lower thermal inertia, which is attributed to higher porosity (60), consistent with their rougher and more crumbly (friable) appearance (Fig. 3, C and D). Conversely, the apparently smoother, more consolidated and angular bright boulders (Fig. 3, A and B) tend to have higher thermal inertias, attributed to lower porosity (60). Distinct texture and porosity can follow from differing mineralogy, levels of compaction, or heating and aqueous alteration histories.

The average blue spectrum of Bennu is consistent with serpentine or magnetite and serpentine, with some carbon (fig. S4, C and D). This composition is similar to that of Bennu's presumed meteorite analogs, the CM and CI groups of carbonaceous chondrites, which are dominated by Fe- and Mg-bearing phyllosilicates, respectively. In the most intensely aqueously altered CI chondrites, the Fe is contained in oxides, sulfides, carbonates, and other minor species (61–63).

The tendency of dark boulders to be slightly redder and brighter in the near-UV region (Fig. 2, A to C) may indicate that they have a higher overall abundance of organic molecules compared with their bright counterparts. This composition manifests as lower reflectance and redder slopes in more freshly exposed surfaces, which develop into steeper blue slopes during the early stages of space weathering. This, along with the weak 0.7- μm absorption sometimes present in dark boulders (Fig. 2D), is indicative of a higher proportion of Fe-bearing

phyllosilicates than other boulders and implies that they represent a population of material from Bennu's parent body that has experienced less aqueous alteration. A correlation between lower reflectance and a 0.7- μm absorption band is consistent with moderately altered CM carbonaceous chondrites, which are among the darkest primitive meteorites with a nearly ubiquitous 0.7- μm absorption feature (64).

Some bright boulders contain veins of brighter material, which may be composed of carbonates (9). This implies a level of aqueous alteration that should also lead to the formation of magnetite (65). The lack of a near-UV upturn in carbonate-bearing boulders could signify that other phases influence their spectral characteristics. For example, bright boulders may be dominated by a hydrated mineral that is more absorbing at the longer wavelengths (so the reflectance is bluer).

We expect vein-bearing boulders to contain less organic material if their reduced carbon was oxidized by fluids to form the observed carbonates (66). Thus, if both low albedo and near-UV bluing are caused by the carbonization of organics via space weathering, we expect those spectrophotometric effects to be weaker in the vein-bearing rocks—as is observed (Fig. 2, A and C). Although graphitized carbon could explain the bluing observed in the more recently exposed faces of dark boulders, a UV upturn is also consistent with increasing abundances of magnetite, which has a blue spectral slope shortward of 0.5 μm (65).

The presence of potential carbonate veins (9) in bright boulders signifies a greater degree of aqueous alteration than in their dark counterparts, suggesting an origin within the interior of Benu's parent body. In dark boulders, we do not find a clear spectral signature indicative of their provenance; however, their textures may provide clues. Brecciated boulders typically have a host matrix similar in texture to the dark boulders, including examples with clasts of exogenic pyroxene; this implies that these rocks formed near the surface of Benu's parent body (38). Boulders on Benu thus may have originated from different zones within the parent body. The multimodal distribution of the boulder populations implies that different processes (or extents of processes), such as aqueous alteration and heating, led to their distinct spectrophotometric properties.

Comparison with Ryugu

Asteroid Ryugu, visited by the Hayabusa2 spacecraft, is also a low-albedo, carbonaceous near-Earth asteroid. Although both asteroids are

thought to have come from primitive asteroid families in the inner main belt (67), Ryugu, unlike Benu, appears to have experienced partial dehydration (68). The multiband cameras onboard the two spacecraft use similar photometric filters in the visible wavelengths and thus allow a direct comparison of the spectra from each (26, 31, 68). Figure S13 shows the areal distribution of the reflectance and near-UV to NIR spectral slope of each asteroid. The variation in reflectance on Benu is 1.7 times that on Ryugu, and Benu exhibits a bluer overall color. Though the standard deviations of the spectral slope distributions are similar ($\sigma = 0.039$ and $0.034 \mu\text{m}^{-1}$ for Ryugu and Benu, respectively), the distribution of color differs spatially. Ryugu shows large-scale latitudinal color differences: The latitudinal difference is ~ 0.5 the standard deviation of the global color variation [$>99\%$ confidence with mean difference of 0.52σ between the redder mid-latitudinal and the bluer equatorial regions (27)], which has been attributed to regolith migration from the equator to mid-latitudes

during the spin-down of Ryugu (68). A latitudinal color trend is also observed on Benu, but the difference is small compared with its overall color variation [$>99\%$ confidence with mean difference of 0.21σ (31)]. Benu's slightly bluer equatorial region may indicate the presence of more mature material, which is consistent with its increasing rotation rate and the associated global patterns of mass movement across the asteroid (43).

Unlike Ryugu, color variation on Benu appears to be dominated by heterogeneity at the meter scale, likely driven by boulders. This suggests that the extent of recent large-scale mass wasting on Benu may not have been as widespread as the effect of regolith mixing. Episodes of particle ejection have been observed from Benu's surface (42), redistributing its surface material over shorter time scales than those expected for mass wasting (69). Large-scale latitudinal spatial patterns may have been obliterated by particle ejection events, which overturn $\sim 10^4$ g of surface material per orbit (437 days) (69). Of the material

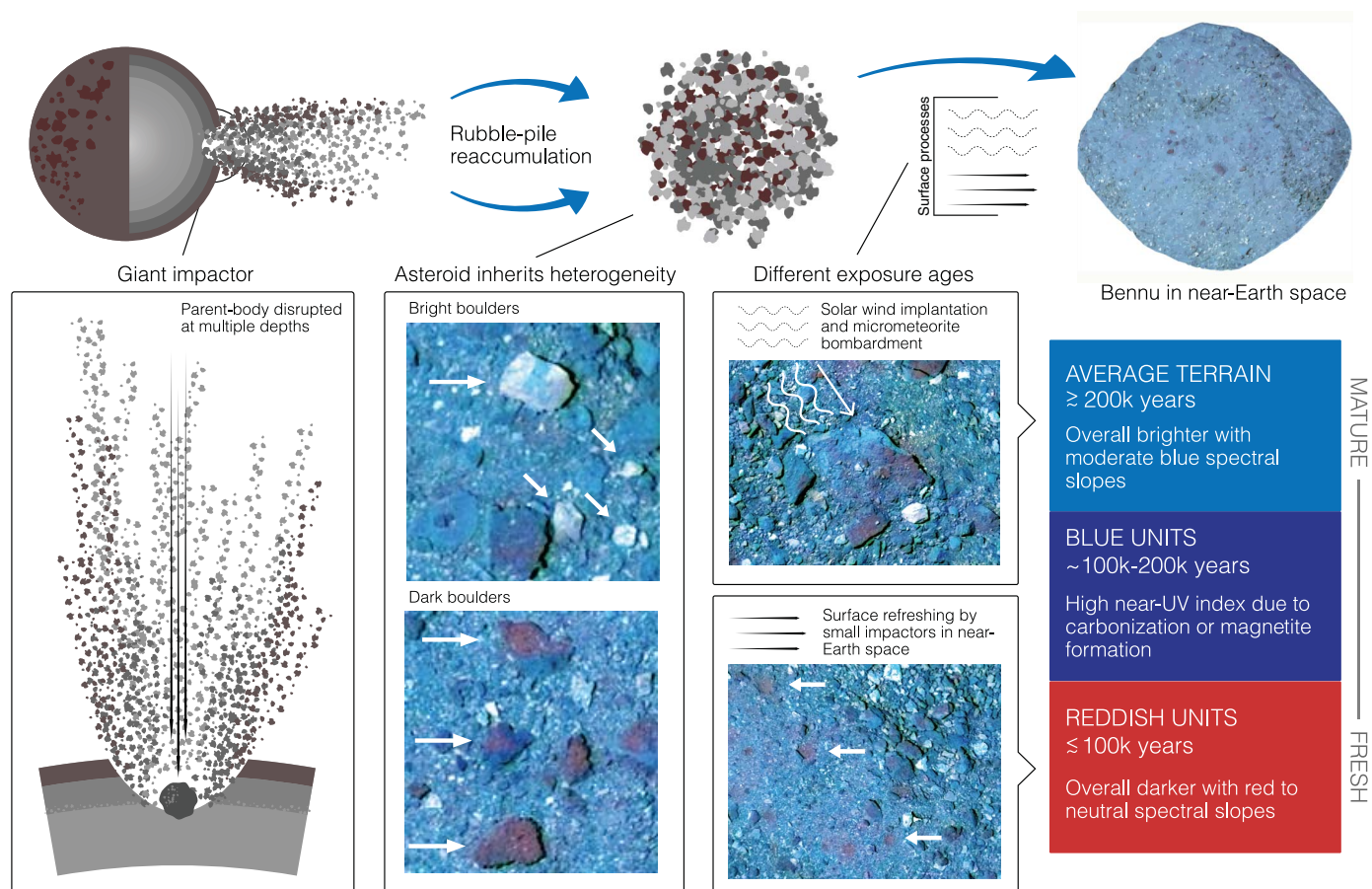


Fig. 7. Proposed model of color and reflectance diversity. We propose that distinct rock types formed at different depths on Benu's parent body. After the parent body was catastrophically disrupted by a giant impactor, Benu accumulated from its debris (4–6) and inherited these distinct materials, leading to the observed heterogeneity among boulders on Benu. Since Benu's

formation and subsequent transit to near-Earth space, it has been altered by exposure to the space environment, which ultimately results in its moderately blue global color. Small near-Earth impactors also continually refresh the surface, creating craters of comparatively fresh, reddish material, which weather toward bluer colors on a time scale of $\sim 10^5$ years.

lofted, most of the mass (70 to 85%) falls back onto Bennu's surface (69).

Young craters on Bennu tend to be redder than the average surface, whereas their young (small) counterparts on Ryugu are bluer than average (70). Because Bennu's global photometric spectrum is bluer overall than Ryugu's, the absolute spectral slopes of the craters on the two asteroids are similar (fig. S14). Hayabusa2 NIR observations showed that the freshly exposed interior of the artificial SCI crater on Ryugu does not exhibit a deep hydration band at 2.7 μm (50), suggesting that Ryugu's dehydration is not a recent event and likely took place on the parent body (71). Although the colors of young craters on Ryugu and Bennu are similar, their hydration properties differ, suggesting that spectral changes in response to space weathering (e.g., bluing versus reddening) can be influenced by initial composition.

Conclusions and prospects for sample return

Bennu's surface is highly diverse, encompassing primitive material potentially from different depths in its parent body (Fig. 7). A smaller proportion consists of exogenic material from another asteroid family delivered in a predispersion impact to the parent body (38). Although Bennu's low average reflectance (0.044) (4, 10) is dominated by the abundant dark boulders and particles formed by their breakdown, the limited latitudinal pattern in the observed heterogeneity indicates a well-mixed combination of disparate materials at spatial scales of 1 to 10 m.

Interboulder variations in reflectance and texture appear to be primordial in origin, but variations in spectral slopes among craters and between individual rock faces appear to be linked to exposure age (Fig. 7). The young age of small reddish craters and the solar orientation of intraboulder color patterns indicate that redder spectra correspond to the most recently exposed surfaces. The underlying composition of dark materials on Bennu, which potentially contain a higher proportion of organic material available for carbonization, likely leads to initial near-UV bluing of any freshly exposed redder materials. Initial bluing could also occur with the production of magnetite in response to space weathering. With age, surface materials brighten and become more neutrally sloped, consistent with Bennu's gently blue average spectral slope. This final stage of space weathering may result from accumulations of nanophase and larger metallic iron particles, which potentially have a brightening and reddening effect in low-reflectance and carbonaceous materials (59). It could also arise from the development of metal-like optical properties in Bennu surface materials due to progressively stronger carbonization effects, which eventually leads to brightening (57).

Our observations suggest that the OSIRIS-REx sample will contain materials with diverse

origins and evolution even from a single location on the asteroid. Both the primary and back-up sample sites, Nightingale (56°N, 42°E) and Osprey (11.5°N, 87.5°E), are situated within small reddish craters. The redder colors and shorter 2.7- μm band minimum positions of materials in these craters imply that they are pristine, have experienced less modification from space weathering than the average Bennu surface, and potentially have a higher proportion of micrometer-scale grains.

REFERENCES AND NOTES

1. D. S. Lauretta *et al.*, OSIRIS-REx: Sample return from asteroid (101955) Bennu. *Space Sci. Rev.* **212**, 925–984 (2017). doi: [10.1007/s11214-017-0405-1](https://doi.org/10.1007/s11214-017-0405-1)
2. B. E. Clark *et al.*, Asteroid (101955) 1999 RQ36: Spectroscopy from 0.4 to 2.4 μm and meteorite analogs. *Icarus* **216**, 462–475 (2011). doi: [10.1016/j.icarus.2011.08.021](https://doi.org/10.1016/j.icarus.2011.08.021)
3. W. F. Bottke *et al.*, In search of the source of asteroid (101955) Bennu: Applications of the stochastic YORP model. *Icarus* **247**, 191–217 (2015). doi: [10.1016/j.icarus.2014.09.046](https://doi.org/10.1016/j.icarus.2014.09.046)
4. D. N. DellaGiustina *et al.*, Properties of rubble-pile asteroid (101955) Bennu from OSIRIS-REx imaging and thermal analysis. *New Astron.* **3**, 341–351 (2019). doi: [10.1038/s41550-019-0731-1](https://doi.org/10.1038/s41550-019-0731-1)
5. D. S. Lauretta *et al.*, The unexpected surface of asteroid (101955) Bennu. *Nature* **568**, 55–60 (2019). doi: [10.1038/s41586-019-1033-6](https://doi.org/10.1038/s41586-019-1033-6); pmid: [30890786](https://pubmed.ncbi.nlm.nih.gov/30890786/)
6. O. S. Barnouin *et al.*, Shape of (101955) Bennu indicative of a rubble pile with internal stiffness. *Nat. Geosci.* **12**, 247–252 (2019). doi: [10.1038/s41561-019-0330-x](https://doi.org/10.1038/s41561-019-0330-x); pmid: [31080497](https://pubmed.ncbi.nlm.nih.gov/31080497/)
7. V. E. Hamilton *et al.*, Evidence for widespread hydrated minerals on asteroid (101955) Bennu. *Nat. Astron.* **3**, 332–340 (2019). doi: [10.1038/s41550-019-0722-2](https://doi.org/10.1038/s41550-019-0722-2); pmid: [31360777](https://pubmed.ncbi.nlm.nih.gov/31360777/)
8. A. A. Simon *et al.*, Widespread carbon-bearing materials on near-Earth asteroid (101955) Bennu. *Science* **370**, eabc3522 (2020). doi: [10.1126/science.abc3522](https://doi.org/10.1126/science.abc3522)
9. H. H. Kaplan *et al.*, Bright carbonate veins on asteroid (101955) Bennu: Implications for aqueous alteration history. *Science* **370**, eabc3557 (2020). doi: [10.1126/science.abc3557](https://doi.org/10.1126/science.abc3557)
10. C. W. Hergenrother *et al.*, The operational environment and rotational acceleration of asteroid (101955) Bennu from OSIRIS-REx observations. *Nat. Commun.* **10**, 1291 (2019). doi: [10.1038/s41467-019-09213-x](https://doi.org/10.1038/s41467-019-09213-x); pmid: [30890725](https://pubmed.ncbi.nlm.nih.gov/30890725/)
11. A. E. Saal, E. H. Hauri, J. A. Van Orman, M. J. Rutherford, Hydrogen isotopes in lunar volcanic glasses and melt inclusions reveal a carbonaceous chondrite heritage. *Science* **340**, 1317–1320 (2013). doi: [10.1126/science.1235142](https://doi.org/10.1126/science.1235142); pmid: [23661641](https://pubmed.ncbi.nlm.nih.gov/23661641/)
12. C. R. Chapman, J. W. Salisbury, Comparisons of meteorite and asteroid spectral reflectivities. *Icarus* **19**, 507–522 (1973). doi: [10.1016/0019-1035\(73\)90078-X](https://doi.org/10.1016/0019-1035(73)90078-X)
13. D. Nesvorný, R. Jedicke, R. J. Whiteley, Ž. Ivezić, Evidence for asteroid space weathering from the Sloan Digital Sky Survey. *Icarus* **173**, 132–152 (2005). doi: [10.1016/j.icarus.2004.07.026](https://doi.org/10.1016/j.icarus.2004.07.026)
14. M. J. Gaffey *et al.*, Mineralogical variations within the S-type asteroid class. *Icarus* **106**, 573–602 (1993). doi: [10.1006/icar.1993.1194](https://doi.org/10.1006/icar.1993.1194)
15. M. Lazzarin *et al.*, Space weathering in the main asteroid belt: The big picture. *Astrophys. J.* **647**, L179–L182 (2006). doi: [10.1086/507448](https://doi.org/10.1086/507448)
16. H. M. Kaluna, J. R. Masiero, K. J. Meech, Space weathering trends among carbonaceous asteroids. *Icarus* **264**, 62–71 (2016). doi: [10.1016/j.icarus.2015.09.007](https://doi.org/10.1016/j.icarus.2015.09.007)
17. R. Brunetto, M. J. Loeffler, D. Nesvorný, S. Sasaki, G. Strazzulla, "Asteroid surface alteration by space weathering processes" in *Asteroids IV*, P. Michel, F. E. DeMeo, W. F. Bottke, Eds. (Univ. of Arizona Press, 2015), pp. 597–616.
18. P. Vernazza *et al.*, Paucity of Tagish Lake-like parent bodies in the Asteroid Belt and among Jupiter Trojans. *Icarus* **225**, 517–525 (2013). doi: [10.1016/j.icarus.2013.04.019](https://doi.org/10.1016/j.icarus.2013.04.019)
19. M. Matsuoka *et al.*, Pulse-laser irradiation experiments of Murchison CM2 chondrite for reproducing space weathering on C-type asteroids. *Icarus* **254**, 135–143 (2015). doi: [10.1016/j.icarus.2015.02.029](https://doi.org/10.1016/j.icarus.2015.02.029)
20. C. Lantz *et al.*, Ion irradiation of carbonaceous chondrites: A new view of space weathering on primitive asteroids. *Icarus* **285**, 43–57 (2017). doi: [10.1016/j.icarus.2016.12.019](https://doi.org/10.1016/j.icarus.2016.12.019)
21. C. Lantz, R. P. Binzel, F. E. DeMeo, Space weathering trends on carbonaceous asteroids: A possible explanation for Bennu's blue slope? *Icarus* **302**, 10–17 (2018). doi: [10.1016/j.icarus.2017.11.010](https://doi.org/10.1016/j.icarus.2017.11.010)
22. M. S. Thompson *et al.*, The effect of progressive space weathering on the organic and inorganic components of a carbonaceous chondrite. *Icarus* **346**, 113775 (2020). doi: [10.1016/j.icarus.2020.113775](https://doi.org/10.1016/j.icarus.2020.113775)
23. H. M. Kaluna, H. A. Ishii, J. P. Bradley, J. J. Gillis-Davis, P. G. Lucey, Simulated space weathering of Fe- and Mg-rich aqueously altered minerals using pulsed laser irradiation. *Icarus* **292**, 245–258 (2017). doi: [10.1016/j.icarus.2016.12.028](https://doi.org/10.1016/j.icarus.2016.12.028)
24. M. S. Thompson, M. J. Loeffler, R. V. Morris, L. P. Keller, R. Christoffersen, Spectral and chemical effects of simulated space weathering of the Murchison CM2 carbonaceous chondrite. *Icarus* **319**, 499–511 (2019). doi: [10.1016/j.icarus.2018.09.022](https://doi.org/10.1016/j.icarus.2018.09.022)
25. D. N. DellaGiustina *et al.*, Overcoming the challenges associated with image-based mapping of small bodies in preparation for the OSIRIS-REx mission to (101955) Bennu. *Earth Space Sci.* **5**, 929–949 (2018). doi: [10.1029/2018EA000382](https://doi.org/10.1029/2018EA000382)
26. B. Rizk *et al.*, OCAMS: The OSIRIS-REx camera suite. *Space Sci. Rev.* **214**, 26 (2018). doi: [10.1007/s11214-017-0460-7](https://doi.org/10.1007/s11214-017-0460-7)
27. B. Zellner, D. J. Tholen, E. F. Tedesco, The eight-color asteroid survey: Results for 589 minor planets. *Icarus* **61**, 355–416 (1985). doi: [10.1016/0019-1035\(85\)90133-2](https://doi.org/10.1016/0019-1035(85)90133-2)
28. F. E. DeMeo, B. Carry, The taxonomic distribution of asteroids from multi-filter all-sky photometric surveys. *Icarus* **226**, 723–741 (2013). doi: [10.1016/j.icarus.2013.06.027](https://doi.org/10.1016/j.icarus.2013.06.027)
29. D. R. Golish *et al.*, Ground and In-Flight Calibration of the OSIRIS-REx Camera Suite. *Space Sci. Rev.* **216**, 12 (2020). doi: [10.1007/s11214-019-0626-6](https://doi.org/10.1007/s11214-019-0626-6); pmid: [32025061](https://pubmed.ncbi.nlm.nih.gov/32025061/)
30. D. R. Golish *et al.*, Disk-resolved photometric modeling and properties of asteroid (101955) Bennu. *Icarus* **10.1016/j.icarus.2020.113724** (2020). doi: [10.1016/j.icarus.2020.113724](https://doi.org/10.1016/j.icarus.2020.113724)
31. Materials and methods are available as supplementary materials.
32. C. A. Bennett *et al.*, A high-resolution global basemap of (101955) Bennu. *Icarus* **10.1016/j.icarus.2020.113690** (2020). doi: [10.1016/j.icarus.2020.113690](https://doi.org/10.1016/j.icarus.2020.113690)
33. K. J. Walsh *et al.*, Craters, boulders and regolith of (101955) Bennu indicative of an old and dynamic surface. *Nat. Geosci.* **12**, 242–246 (2019). doi: [10.1038/s41561-019-0326-6](https://doi.org/10.1038/s41561-019-0326-6)
34. M. G. Daly *et al.*, The OSIRIS-REx laser altimeter (OLA) investigation and instrument. *Space Sci. Rev.* **212**, 899–924 (2017). doi: [10.1007/s11214-017-0375-3](https://doi.org/10.1007/s11214-017-0375-3)
35. O. S. Barnouin *et al.*, Digital terrain mapping by the OSIRIS-REx mission. *Planet. Space Sci.* **180**, 104764 (2020). doi: [10.1016/j.pss.2019.104764](https://doi.org/10.1016/j.pss.2019.104764)
36. M. G. Daly *et al.*, Hemispherical differences in the shape and topography of asteroid (101955) Bennu. *Sci. Adv.* **6**, eabd3649 (2020). doi: [10.1126/sciadv.abd3649](https://doi.org/10.1126/sciadv.abd3649)
37. D. C. Reuter *et al.*, The OSIRIS-REx Visible and InfraRed Spectrometer (OVIRS): Spectral maps of the asteroid Bennu. *Space Sci. Rev.* **214**, 54 (2018). doi: [10.1007/s11214-018-0482-9](https://doi.org/10.1007/s11214-018-0482-9)
38. D. N. DellaGiustina *et al.*, Exogenic Basalt on Asteroid (101955) Bennu. *New Astron.* **10.1038/s41550-020-1195-z** (2020). doi: [10.1038/s41550-020-1195-z](https://doi.org/10.1038/s41550-020-1195-z)
39. J. L. Molero *et al.*, In situ evidence of thermally induced rock breakdown widespread on Bennu's surface. *Nat. Commun.* **11**, 2913 (2020). doi: [10.1038/s41467-020-16528-7](https://doi.org/10.1038/s41467-020-16528-7); pmid: [32518333](https://pubmed.ncbi.nlm.nih.gov/32518333/)
40. C. T. Russell *et al.*, Dawn at Vesta: Testing the protoplanetary paradigm. *Science* **336**, 684–686 (2012). doi: [10.1126/science.1219381](https://doi.org/10.1126/science.1219381); pmid: [22582253](https://pubmed.ncbi.nlm.nih.gov/22582253/)
41. F. Vilas, A cheaper, faster, better way to detect water of hydration on Solar System bodies. *Icarus* **111**, 456–467 (1994). doi: [10.1006/icar.1994.1156](https://doi.org/10.1006/icar.1994.1156)
42. D. S. Lauretta *et al.*, Episodes of particle ejection from the surface of the active asteroid (101955) Bennu. *Science* **366**, eaay3544 (2019). doi: [10.1126/science.aay3544](https://doi.org/10.1126/science.aay3544); pmid: [31806784](https://pubmed.ncbi.nlm.nih.gov/31806784/)
43. E. R. Javin *et al.*, Global patterns of mass movement on asteroid (101955) Bennu. *J. Geophys. Res. Planets* **125**, e2020JE006475 (2020). doi: [10.1029/2020JE006475](https://doi.org/10.1029/2020JE006475)
44. A. R. Hendrix, F. Vilas, C-complex asteroids: UV-visible spectral characteristics and implications for space weathering effects. *Geophys. Res. Lett.* **46**, 14307–14317 (2019). doi: [10.1029/2019GL085883](https://doi.org/10.1029/2019GL085883)
45. W. F. Bottke Jr. *et al.*, Linking the collisional history of the main asteroid belt to its dynamical excitation and depletion. *Icarus* **179**, 63–94 (2005). doi: [10.1016/j.icarus.2005.05.017](https://doi.org/10.1016/j.icarus.2005.05.017)

46. S. Marchi *et al.*, The cratering history of asteroid (2867) Steins. *Planet. Space Sci.* **58**, 1116–1123 (2010). doi: [10.1016/j.pss.2010.03.017](https://doi.org/10.1016/j.pss.2010.03.017)
47. P. Brown, R. E. Spalding, D. O. ReVelle, E. Tagliaferri, S. P. Worden, The flux of small near-Earth objects colliding with the Earth. *Nature* **420**, 294–296 (2002). doi: [10.1038/nature01238](https://doi.org/10.1038/nature01238); pmid: [12447433](https://pubmed.ncbi.nlm.nih.gov/12447433/)
48. A. W. Harris, G. D'Abamo, The population of near-Earth asteroids. *Icarus* **257**, 302–312 (2015). doi: [10.1016/j.icarus.2015.05.004](https://doi.org/10.1016/j.icarus.2015.05.004)
49. S. Marchi, C. R. Chapman, O. S. Barnouin, J. E. Richardson, J.-B. Vincent, "Cratering on asteroids" in *Asteroids IV*, P. Michel, F. E. DeMeo, W. F. Bottke, Eds. (Univ. of Arizona Press, 2015), pp. 725–744.
50. M. Arakawa *et al.*, An artificial impact on the asteroid (162173) Ryugu formed a crater in the gravity-dominated regime. *Science* **368**, 67–71 (2020). doi: [10.1126/science.aaz1701](https://doi.org/10.1126/science.aaz1701); pmid: [32193363](https://pubmed.ncbi.nlm.nih.gov/32193363/)
51. E. Tatsumi, S. Sugita, Cratering efficiency on coarse-grain targets: Implications for the dynamical evolution of asteroid 25143 Itokawa. *Icarus* **300**, 227–248 (2018). doi: [10.1016/j.icarus.2017.09.004](https://doi.org/10.1016/j.icarus.2017.09.004)
52. L. P. Keller, E. L. Berger, A transmission electron microscope study of Itokawa regolith grains. *Earth Planets Space* **66**, 71 (2014). doi: [10.1186/1880-5981-66-71](https://doi.org/10.1186/1880-5981-66-71)
53. T. V. Johnson, F. P. Fanale, Optical properties of carbonaceous chondrites and their relationship to asteroids. *J. Geophys. Res.* **78**, 8507–8518 (1973). doi: [10.1029/JB078i035p08507](https://doi.org/10.1029/JB078i035p08507)
54. C. M. Hartzell, Dynamics of 2D electrostatic dust levitation at asteroids. *Icarus* **333**, 234–242 (2019). doi: [10.1016/j.icarus.2019.05.013](https://doi.org/10.1016/j.icarus.2019.05.013)
55. M. C. Nolan *et al.*, Detection of rotational acceleration of Benu using HST light curve observations. *Geophys. Res. Lett.* **46**, 1956–1962 (2019). doi: [10.1029/2018GL080658](https://doi.org/10.1029/2018GL080658)
56. W. F. Bottke *et al.*, Meteoroid Impacts as a Source of Benu's Particle Ejection Events. *J. Geophys. Res. Planets* **125**, e2019JE006282 (2020). doi: [10.1029/2019JE006282](https://doi.org/10.1029/2019JE006282)
57. L. Moroz *et al.*, Optical alteration of complex organics induced by ion irradiation: 1. Laboratory experiments suggest unusual space weathering trend. *Icarus* **170**, 214–228 (2004). doi: [10.1016/j.icarus.2004.02.003](https://doi.org/10.1016/j.icarus.2004.02.003)
58. B. Hapke, Space weathering from Mercury to the asteroid belt. *J. Geophys. Res. Planets* **106**, 10039–10073 (2001). doi: [10.1029/2000JE001338](https://doi.org/10.1029/2000JE001338)
59. A. S. Rivkin, E. S. Howell, F. Vilas, L. A. Lebofsky, "Hydrated minerals on asteroids: The astronomical record" in *Asteroids III*, W. F. Bottke, A. Cellino, P. Paolicchi, R. P. Binzel (Univ. of Arizona Press, 2002), pp. 235–253.
60. B. Rozitis *et al.*, Asteroid (101955) Benu's weak boulders and thermally anomalous equator. *Sci. Adv.* **6**, eabc3699 (2020). doi: [10.1126/sciadv.abc3699](https://doi.org/10.1126/sciadv.abc3699)
61. C. A. Johnson, M. Prinz, Carbonate compositions in CM and CI chondrites and implications for aqueous alteration. *Geochim. Cosmochim. Acta* **57**, 2843–2852 (1993). doi: [10.1016/0016-7037\(93\)90393-B](https://doi.org/10.1016/0016-7037(93)90393-B)
62. A. Morlok *et al.*, Brecciation and chemical heterogeneities of CI chondrites. *Geochim. Cosmochim. Acta* **70**, 5371–5394 (2006). doi: [10.1016/j.gca.2006.08.007](https://doi.org/10.1016/j.gca.2006.08.007)
63. J. Alving, M. Patzek, A. Bischoff, Modal abundances of coarse-grained (> 5 μm) components within CI-chondrites and their individual clasts—Mixing of various lithologies on the CI parent body(ies). *Geochemistry* **79**, 125532 (2019). doi: [10.1016/j.chemer.2019.08.004](https://doi.org/10.1016/j.chemer.2019.08.004)
64. E. A. Cloutis *et al.*, Spectral reflectance properties of HED meteorites CM2 carbonaceous chondrites: Comparison to HED grain size and compositional variations and implications for the nature of low-albedo features on Asteroid 4 Vesta. *Icarus* **223**, 850–877 (2013). doi: [10.1016/j.icarus.2013.02.003](https://doi.org/10.1016/j.icarus.2013.02.003)
65. M. R. Izawa *et al.*, Spectral reflectance properties of magnetites: Implications for remote sensing. *Icarus* **319**, 525–539 (2019). doi: [10.1016/j.icarus.2018.10.002](https://doi.org/10.1016/j.icarus.2018.10.002)
66. C. M. Alexander, R. Bowden, M. L. Fogel, K. T. Howard, Carbonate abundances and isotopic compositions in chondrites. *Meteorit. Planet. Sci.* **50**, 810–833 (2015). doi: [10.1111/maps.12410](https://doi.org/10.1111/maps.12410)
67. J. de León *et al.*, Expected spectral characteristics of (101955) Benu and (162173) Ryugu, targets of the OSIRIS-REx and Hayabusa2 missions. *Icarus* **313**, 25–37 (2018). doi: [10.1016/j.icarus.2018.05.009](https://doi.org/10.1016/j.icarus.2018.05.009)
68. S. Sugita *et al.*, The geomorphology, color, and thermal properties of Ryugu: Implications for parent-body processes. *Science* **364**, 252 (2019). doi: [10.1126/science.aaw0422](https://doi.org/10.1126/science.aaw0422); pmid: [30890587](https://pubmed.ncbi.nlm.nih.gov/30890587/)
69. C. W. Hergenrother *et al.*, Photometry of particles ejected from active asteroid (101955) Benu. *J. Geophys. Res. Planets* **125**, e2020JE006381 (2020). doi: [10.1029/2020JE006381](https://doi.org/10.1029/2020JE006381)
70. T. Morota *et al.*, Sample collection from asteroid (162173) Ryugu by Hayabusa2: Implications for surface evolution. *Science* **368**, 654–659 (2020). doi: [10.1126/science.aaz6306](https://doi.org/10.1126/science.aaz6306); pmid: [32381723](https://pubmed.ncbi.nlm.nih.gov/32381723/)
71. K. Kitazato *et al.*, "Asteroid 162173 Ryugu: Surface composition as observed by Hayabusa2/NIRS3," in *EPSC Abstracts*, vol. 13, EPSC-DPS2019-1376-1 (EPSC, 2019); <https://meetingorganizer.copernicus.org/EPSC-DPS2019/EPSC-DPS2019-1376-1.pdf>.
72. D. DellaGiustina, Maps_DellaGiustina_et_al_Science_2020_abc3660, Version 2, figshare (2020); <https://doi.org/10.6084/m9.figshare.12996494>.

ACKNOWLEDGMENTS

We thank the entire OSIRIS-REx Team for making the encounter with Benu possible. For fig. S5, we thank D. Rachford and IMV Minerals for providing the SAPI05 sample and S. Mertzman of Franklin and Marshall College for confirming the composition of the SAPI05 sample. We thank R. J. Burmeister for assistance with graphic design of the figures. **Funding:** D.N.D., K.N.B., K.J.W., P.H.S., D.R.G., E.B.B., R.L.B., T.B., H.C., B.E.C., A.S., K.B., C.A.B., W.F.B., J.I.B., H.C.C., C.Y.D.D., K.L.E., V.E.H., C.W.H., E.S.H., E.R.J., H.H.K., L.L.C., L.L., J.Y.L., M.C.N., J.N., N.A.P., B.Ri., A.J.R., N.K.S., A.A.S., C.W.V.W., and D.S.L. were supported by NASA under contract NNM10AA11C issued through the New Frontiers Program. J.L.M. and D.T. were supported by the NASA OSIRIS-REx Participating Scientist Program 80NSC18K0230. A.R.H. was supported by the NASA Solar System Exploration Research Virtual Institute 2016 (SSERV16) Cooperative Agreement TREX NNH16ZDA001N. E.T., K.Y., and S.S. were supported by the Japanese Society for Promotion of Science (JSPS) Core-to-Core program "International Planetary Network." The OLA instrument and funding for M.G.D. and M.M.A.A. and support for A.P., E.A.C., and D.M.A., were provided by the Canadian Space Agency. J.L.R., E.T., M.Po., and J.d.L. acknowledge support from the project AYA2017-89090-P of the Spanish MINECO and from the project

ProlD2017010112 under the OP-ERDF-ESF and the Canarian Agency for Research, Innovation and Information Society (ACIISI). M.Pa. was supported by the Italian Space Agency (ASI) under the ASI-INF agreement no. 2017-37-H.O. C.A., J.D.P.D., M.A.B., M.D., P.H.H., P.M., and S.F. acknowledge support from the French space agency CNES. P.M. also acknowledges support from the European Union's Horizon 2020 research and innovation program under grant agreement no. 870377 (project NEO-MAPP). C.A. was also supported by the French National Research Agency under the project "Investissements d'Avenir" UCAJEDI (ANR-15-IDEX-01). C.A. and M.D. also acknowledge support from the ANR "ORIGINS" (ANR-18-CE31-0014). B.Ro. acknowledges support from the Royal Astronomical Society (RAS) and the UK Science and Technology Facilities Council (STFC). **Author contributions:** D.N.D., P.H.S., D.R.G., B.Ri., and C.Y.D.D. planned and calibrated OCAMS MapCam and PolyCam observations of Benu. D.N.D., K.N.B., P.H.S., D.R.G., T.L.B., K.J.B., C.A.B., J.I.B., K.L.E., L.L.C., and N.K.S. created MapCam color parameter maps of Benu. K.N.B., E.B.B., J.N., D.T., and R.B.V.A. mapped craters and boulders on Benu. D.N.D., K.N.B., K.J.W., P.H.S., E.B.B., R.-L.B., M.M.A.A., and W.F.B. performed the statistical analysis of boulder and craters colors. E.T., K.Y., and S.S. performed comparisons of the colors of Benu and Ryugu. E.A.C., B.E.C., A.R.H., A.S., D.M.A., and A.P. performed comparisons between Benu colors and laboratory spectra. J.D.P.D., A.A.S., H.H.K., and V.E.H. processed and interpreted OVIRS observations of Benu. M.G.D., M.M.A.A., R.-L.B., and N.A.P. processed and interpreted OLA observations of Benu. D.N.D., P.H.S., D.R.G., K.J.W., R.-L.B., H.C., C.A., M.A.B., H.C.C.Jr., M.D., J.d.L., S.F., V.E.H., P.H.H., C.W.H., E.S.H., E.R.J., H.H.K., L.F.L., J.Y.L., P.M., J.L.M., M.C.N., M.Po., A.J.R., B.Ro., A.A.S., D.T., and C.W.V.W. contributed to the conceptualization and writing. D.S.L. contributed to the conceptualization and leads the OSIRIS-REx mission. **Competing interests:** We declare no competing interests. **Data and materials availability:** Full-resolution MapCam color maps of Benu, as shown in Fig. 1 and fig. S5, are available at figshare (72). Catalogs of the boulders and craters we used are available as data S1 and S2, respectively. The stereophotoclinometry v28 and OLA v20 shape models of Benu are available through the Small Body Mapping Tool at <http://sbmt.jhuapl.edu/>. Detailed Survey OCAMS (MapCam and PolyCam), Orbital B OLA, and Detailed Survey and Recon A OVIRS data are available via the Planetary Data System (PDS) at: https://sbnarchive.psi.edu/pds4/orex/orex.ocams/data_calibrated/detailed_survey/, https://sbnarchive.psi.edu/pds4/orex/orex.ocams/data_calibrated/recon/, https://sbnarchive.psi.edu/pds4/orex/orex.ola/data_calibrated/orbit_b/, https://sbnarchive.psi.edu/pds4/orex/orex.ovirs/data_calibrated/detailed_survey/, and https://sbnarchive.psi.edu/pds4/orex/certified/orex.ovirs/data_calibrated/recon/, respectively. The list of MapCam images used to produce the maps in Fig. 1 is available as data S3.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6517/eabc3660/suppl/DC1
Materials and Methods
Figs. S1 to S16
Tables S1 and S2
References (73–91)
Data S1 to S3

21 April 2020; accepted 24 September 2020
Published online 8 October 2020
[10.1126/science.abc3660](https://doi.org/10.1126/science.abc3660)

RESEARCH ARTICLE SUMMARY

ASTEROIDS

Widespread carbon-bearing materials on near-Earth asteroid (101955) Bennu

Amy A. Simon*, Hannah H. Kaplan, Victoria E. Hamilton, Dante S. Lauretta, Humberto Campins, Joshua P. Emery, M. Antonietta Barucci, Daniella N. DellaGiustina, Dennis C. Reuter, Scott A. Sandford, Dathon R. Golish, Lucy F. Lim, Andrew Ryan, Benjamin Rozitis, Carina A. Bennett

INTRODUCTION: Owing to their low reflectance and spectral similarity to primitive carbonaceous chondrite meteorites, C-complex asteroids are thought to contain carbon-bearing material. The OSIRIS-REx (Origins, Spectral Interpretation, Resource Identification, and Security–Regolith Explorer) spacecraft is designed to return a sample of carbonaceous material from the near-Earth C-complex asteroid (101955) Bennu. The selection of a suitable sample site necessitated global mapping and characterization of Bennu's surface. Spatially resolved spectral mapping can determine the surface properties and composition of Bennu. It also provides context for both the sample that will be returned and the interpretation of unresolved observations of other dark asteroids.

RATIONALE: We used data acquired by the OSIRIS-REx Visible and InfraRed Spectrometer (OVIRS), a point spectrometer that covers the wavelength range from 0.4 to 4.3 μm , to map the physical and compositional characteristics of Bennu's surface. These data allow us to search for spectral signatures of carbon bearing materials on Bennu. The 3.4- μm region is sensitive to carbonate or organic materials, which produce absorption bands at this wavelength because of either CO_3^{2-} stretching and

vibration or C-H stretching. OVIRS mapping provided global coverage of Bennu at $\sim 600\text{-m}^2$ areal resolution at several local solar times. Using the data with the highest solar illumination ($\sim 9^\circ$ phase, 12:30 p.m. local solar time), we mapped the depth of the 3.4- μm absorption band, peak temperature, 0.55- μm brightness, spectral slope from 0.5 to 1.5 μm , and the distribution of the 2.74- μm absorption band of hydrated minerals, which was previously detected in unresolved observations.

RESULTS: The 3.4- μm absorption band, indicative of carbon-bearing materials, is detected over all of Bennu's surface with band depths of a few percent. The band shape varies with surface location and spans the range of 3.4- μm band shapes seen on other dark C-complex asteroids. The differing band shapes persist at higher areal resolution (60 m^2) and at several phase angles. The spectra collected at 60 m^2 show that the deepest bands occur over distinct boulders.

The distribution of the 3.4- μm band on Bennu's surface does not correlate with the distributions of temperature, brightness, spectral slope, or the 2.74- μm absorption band, although some of these features correlate weakly with each other. At low phase angles, the darkest areas ($\sim 3\%$ reflectance at 0.55 μm) are correlated

with the hottest surface temperatures ($\sim 350\text{ K}$), with a Spearman's rank correlation coefficient, r , of 0.65.

The absorption feature at 2.74 μm , indicative of hydrated phyllosilicates, is globally present, with band depths of 12 to 17% that correlate with surface temperature and latitude ($|r| = 0.76$ and 0.58 , respectively). When the temperature trend is removed, the correlation of hydrated phyllosilicates with latitude is weaker ($|r| = 0.48$). In OVIRS data, Bennu's global surface has an overall blue (negative) spectral slope from 0.5 to 1.5 μm , with some boulders and craters that are redder (less negative) than average, consistent with results from multispectral imaging. Some of the darkest material is spectrally blue, whereas some is spectrally red, indicating local differences in composition, space weathering, and/or particle size.

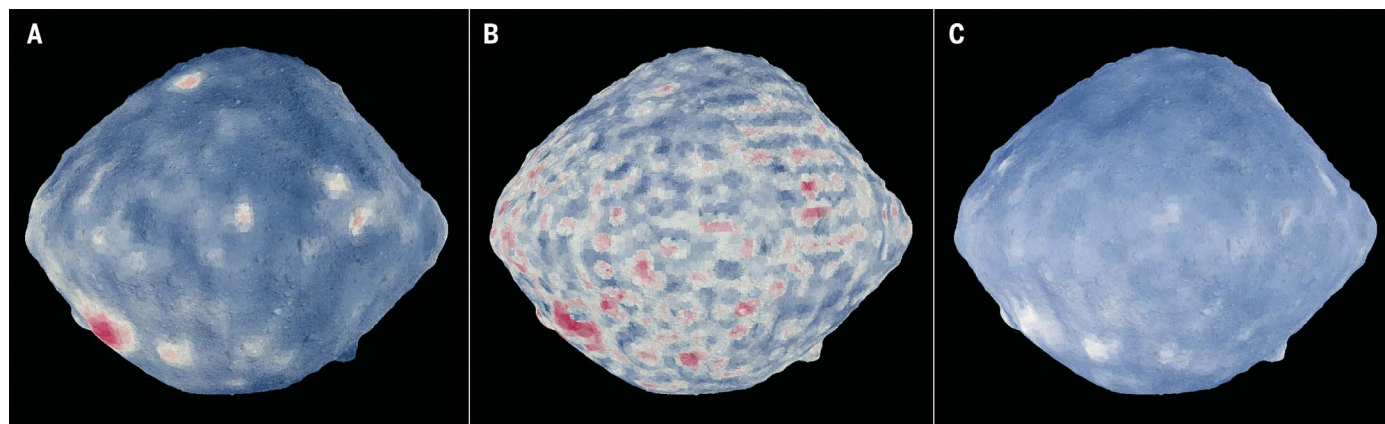
CONCLUSION: The variation in the shape of the 3.4- μm band indicates a mix of organics and carbonates on Bennu's surface, likely inherited from the collisional disruption of its parent asteroid. To retain a widespread 3.4- μm organic feature, most of the material on Bennu's surface could not have been exposed to the space environment for more than a few million years. The samples returned to Earth by the OSIRIS-REx spacecraft should contain ample amounts of these materials, regardless of sampling location. Variable 3.4- μm band depths over individual boulders may be due to compositional differences or to exposure of fresh material by means of thermally driven fracturing. ■

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

*Corresponding author. Email: amy.simon@nasa.gov

Cite this article as A. A. Simon *et al.*, *Science* **370**, eabc3522 (2020). DOI: 10.1126/science.abc3522

S READ THE FULL ARTICLE AT
<https://doi.org/10.1126/science.abc3522>



Spectral variations on Bennu's 60°E hemisphere. (A) Visible to near-infrared (0.5 to 1.5 μm) slope. Blue denotes more steeply negative slopes (decreasing brightness with increasing wavelength); red denotes shallower slopes. (B) Band area at 3.4 μm , indicative of carbon-bearing materials. Blue indicates smaller band areas; red, larger band areas. (C) Band depth at 2.74 μm , indicative of hydrated phyllosilicates. White indicates shallower bands; blue, deeper bands.

RESEARCH ARTICLE

ASTEROIDS

Widespread carbon-bearing materials on near-Earth asteroid (101955) Bennu

Amy A. Simon^{1*}, Hannah H. Kaplan², Victoria E. Hamilton², Dante S. Lauretta³, Humberto Campins⁴, Joshua P. Emery⁵, M. Antonietta Barucci⁶, Daniella N. DellaGiustina³, Dennis C. Reuter¹, Scott A. Sandford⁷, Dathon R. Golish³, Lucy F. Lim¹, Andrew Ryan³, Benjamin Rozitis⁸, Carina A. Bennett³

Asteroid (101955) Bennu is a dark asteroid on an Earth-crossing orbit that is thought to have assembled from the fragments of an ancient collision. We use spatially resolved visible and near-infrared spectra of Bennu to investigate its surface properties and composition. In addition to a hydrated phyllosilicate band, we detect a ubiquitous 3.4-micrometer absorption feature, which we attribute to a mix of organic and carbonate materials. The shape and depth of this absorption feature vary across Bennu's surface, spanning the range seen among similar main-belt asteroids. The distribution of the absorption feature does not correlate with temperature, reflectance, spectral slope, or hydrated minerals, although some of those characteristics correlate with each other. The deepest 3.4-micrometer absorptions occur on individual boulders. The variations may be due to differences in abundance, recent exposure, or space weathering.

Asteroid (101955) Bennu is a dark near-Earth asteroid averaging 4.4% reflectance (1), thought to be a collisional remnant of a much larger object in the main asteroid belt (2). Bennu is classified as a C-complex asteroid—a body with low reflectance and spectral similarities to carbonaceous chondrite meteorites—so is expected to have high carbon content (3). Specifically, Bennu belongs to a subgroup of the C-complex, the B-types, which have bluer spectra (4) and are thought to have silicates, hydrated minerals, organic molecules, magnetite, and sulfides on their surfaces (5).

Launched in 2016, the OSIRIS-REx (Origins, Spectral Interpretation, Resource Identification, and Security–Regolith Explorer) spacecraft is designed to return a sample of carbonaceous material from Bennu to Earth for laboratory analysis (6). One of the mission objectives is to compare ground-based Bennu data with the resolved surface to help interpret unresolved observations of other asteroids (6). Detailed spatial and spectral maps will also be required to tie the returned sample to its local geologic context.

The B-type asteroids are subdivided according to their visible to near-infrared (VIS-NIR)

spectral slopes, with some bluer objects similar to (2) Pallas, some redder objects similar to (24) Themis, and others in the continuum of spectral slopes between them (7). Each of these VIS-NIR spectral slopes can be matched with analogous carbonaceous chondrite meteorites, which show various degrees of aqueous and thermal alteration (5). On the basis of these meteorite comparisons, it has been proposed that the redder-sloped B-types are the most water-rich, similar to the CM class of chondrites, and the bluer-sloped B-types are dry and have been heated or thermally metamorphosed, similar to the CK4 class of chondrites (5). Before the OSIRIS-REx spacecraft's arrival, one of the closest spectral matches to Bennu's blue spectral slope (at <1.5 μm) was a carbonaceous chondrite that shows signs of having been aqueously altered then heated and dehydrated (8). However, global spectra acquired during the spacecraft's approach to Bennu showed a 2.7- μm absorption band, indicating hydrated minerals, with a shape more similar to that of the aqueously altered CM chondrites (9).

Bennu is thought to be a member of either the (142) Polana or the (495) Eulalia collisional family, groups of asteroids that appear to originate from a common source in the main asteroid belt between Mars and Jupiter (10, 11). Both Polana and Eulalia reside in the main belt and have spectra that indicate that they are members of the C-complex (12). The Hayabusa2 spacecraft has visited and sampled the near-Earth asteroid (162173) Ryugu, which is also likely to be a member of the Polana or Eulalia family, but is classified as a near-Earth Cb-type asteroid, a different spectral subgroup of the C-complex (13). Spatially resolved spectra from

Hayabusa2 indicate that Ryugu has a weaker global 2.7- μm hydration feature than Bennu. On Ryugu, the band varies only slightly (7 to 10% band depth) across the surface, which is consistent with thermal processing (14). Other asteroid families, such as the (24) Themis family, show spectral diversity among their members and evidence of material separation, possibly into rock and ice—likely enough to produce aqueous alteration (15). Similar processing may have occurred in the parent body of Bennu. Bennu is a rubble-pile asteroid, an agglomeration of collision fragments, that was ejected from the main asteroid belt and migrated to its final near-Earth orbit. Global maps of Bennu's spectral characteristics are required to understand its relationship to Ryugu and their parent population and to the broader classes of B- and C-type asteroids.

In 2019, the OSIRIS-REx mission completed a campaign of close-proximity observation of Bennu to determine suitable sites for sample collection, which occurred 20 October 2020 (6). The OSIRIS-REx Visible and InfraRed Spectrometer (OVIRS) acquired global surface data from 0.4 to 4.3 μm at multiple phase angles (16). Visible wavelengths are used to determine the overall color variations across Bennu's surface and its mineral content, whereas the NIR is sensitive to surface temperature and compositional variations. The OVIRS NIR coverage around 2.7 μm has been used to detect the global presence of hydrated minerals (9), and the region from 3.2 to 3.6 μm is sensitive to carbonate (CO_3^{2-} stretching or vibration) and organic (C-H stretching) absorption bands.

In this study, we investigated the spectrum of Bennu using OVIRS observations and searched for evidence of absorption bands attributable to carbon-bearing materials. Our goal was to compare the shape of any carbon-bearing absorption features with spectra of other C-complex asteroids and determine the surface distribution of carbon-bearing materials.

Observations and mapping

We used data from the Equatorial Stations subphase (25 April to 6 June 2019) of the OSIRIS-REx Detailed Survey phase, which ran from 7 March through 6 June 2019. The subphase was composed of data collection from seven spacecraft stations, each positioned close to the equator at a different local solar time and ~5 km above the surface (6). As OVIRS is a point spectrometer, with a circular 4-mrad field of view, each spectrum covers all wavelengths simultaneously for a spot on the surface (16). At each local solar time, the spacecraft nodded along lines of constant longitude at ~2 mrad s^{-1} to achieve nearly global OVIRS coverage as the asteroid rotated below. Although OVIRS observed most latitudes and longitudes at each station, the individual spots were not

¹Solar System Exploration Division, NASA Goddard Space Flight Center, Greenbelt, MD, USA. ²Southwest Research Institute, Boulder, CO, USA. ³Lunar and Planetary Laboratory, University of Arizona, Tucson, AZ, USA. ⁴Department of Physics, University of Central Florida, Orlando, FL, USA.

⁵Department of Astronomy and Planetary Sciences, Northern Arizona University, Flagstaff, AZ, USA. ⁶Laboratoire d'Études Spatiales et d'Instrumentation en Astrophysique, Observatoire de Paris, Université Paris Sciences et Lettres, Centre National de la Recherche Scientifique, Université de Paris, Sorbonne Université, Meudon, France. ⁷NASA Ames Research Center, Moffett Field, CA, USA. ⁸School of Physical Sciences, The Open University, Milton Keynes, UK.

*Corresponding author. Email: amy.simon@nasa.gov

centered at the exact same locations from station to station. The spot spatial resolution was ~ 20 m cross-track and 30 m along-track at the equator. Surface projection increases the along-track spot size at higher latitudes, and local surface slopes also cause the resolution to vary. From these data, we produced global OVIRS maps at $\sim 600\text{-m}^2$ equatorial areal resolution from the stations at 10:00 a.m., 12:30 p.m., and 3:00 p.m. local solar time, as these have the highest available solar illumination and maximum surface temperatures.

From 4 to 27 October 2019, the spacecraft conducted lower-altitude reconnaissance flyovers of four candidate sample sites and surrounding context regions, scanning each area of interest at a range of 1 to 1.3 km (6). Reconnaissance flyovers occurred when Bennu was at a greater distance from the Sun than during the Detailed Survey, and therefore solar illumination was lower; longer exposure times were used to compensate. The resulting OVIRS spatial resolution was ~ 4 to 5 m cross-track and ~ 7 to 10 m along-track, plus the surface projection effects at the higher latitudes. We produced spectral maps of these smaller regions at $\sim 60\text{-m}^2$ areal resolution.

Globally averaged spectra

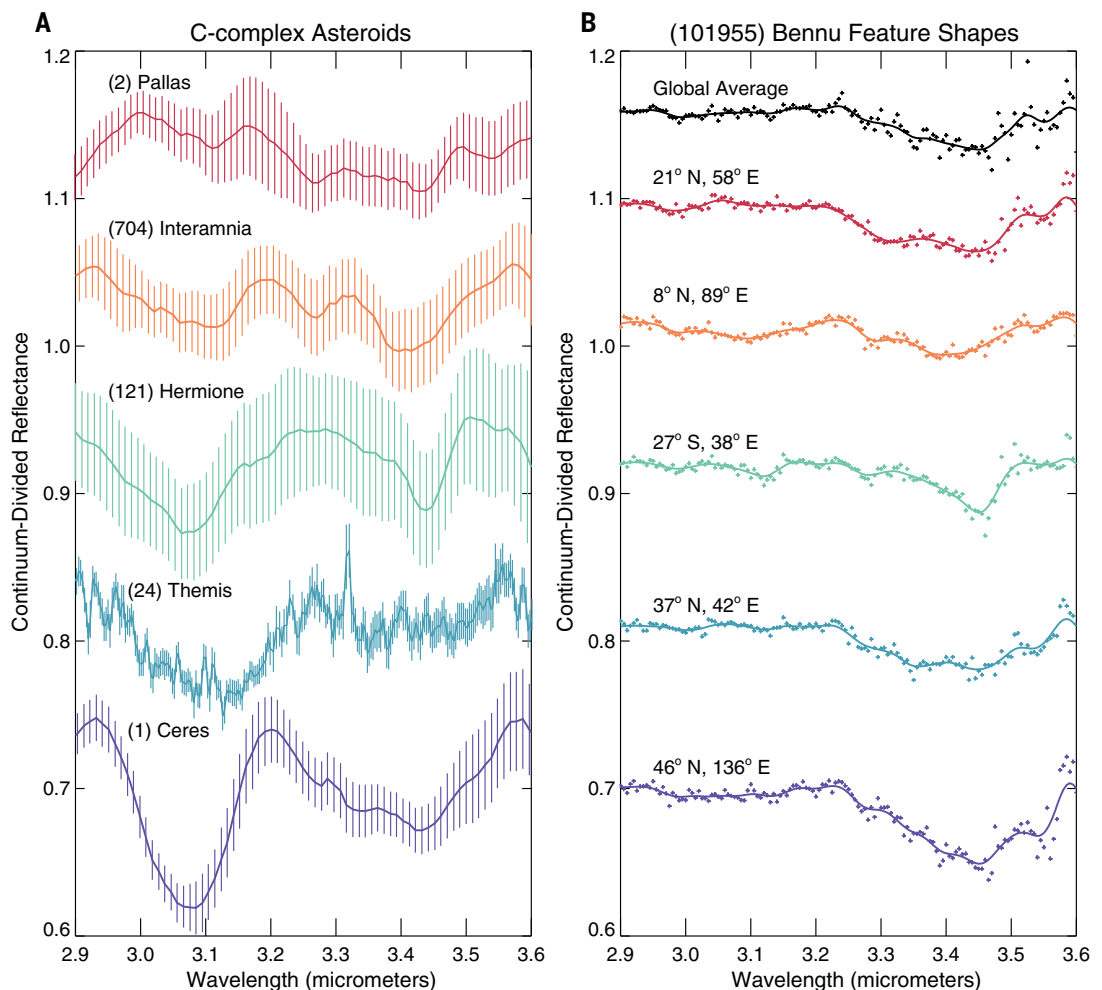
Global spectra were calculated for each Detailed Survey equatorial spectral mapping station by averaging all OVIRS spots that were fully on the surface; there are 6189, 7193, and 6877 spectra for 10:00 a.m., 12:30 p.m., and 3:00 p.m., respectively. Although not identical to a full-disk spectrum, because of the fixed local time, the averages include emission angles from latitudes all the way to the poles. In the 12:30 p.m. data, Bennu shows a weak global absorption feature from 3.2 to 3.6 μm (hereafter, the 3.4- μm feature), which we attribute to carbonate and organic materials (Fig. 1B). This absorption feature was not detectable in full-disk, unresolved OVIRS spectra, taken when Bennu only filled $\sim 40\%$ of the field of view and the surface was warmer owing to closer proximity to the Sun (9). Spatially resolved datasets were also acquired in March 2019, very close to Bennu's perihelion. These contain hits of a 3.4- μm absorption feature, but the band depth was likely affected by thermal fill-in, an effect whereby greater thermal emission at high temperatures increases the thermal flux at the band wavelength, thereby weakening absorption features (17). Those spectra were also noisier

at long wavelengths $> 3 \mu\text{m}$, owing to higher instrument detector temperatures.

Bennu's globally averaged 3.4- μm absorption feature, as observed during the Detailed Survey, is distinct from that seen on other C-complex asteroids (Fig. 1). Previous observations of a 3.4- μm feature in other disk-integrated asteroid spectra, including (24) Themis and Themis-like asteroids, have been attributed primarily to aliphatic organic molecules (18–20). However, ground-based observations of (1) Ceres and Ceres-like asteroids have revealed a strong 3.4- μm feature that was instead primarily attributed to carbonates, as it was accompanied by a detection of the 3.9- μm carbonate band (20, 21). Several main-belt asteroids display variability between 3.2 and 3.6 μm , indicating potential compositional variation, although the band variations are uncertain for (704) Interamnia and (121) Hermione (Fig. 1A) (18, 22). However, in individual spot spectra, Bennu has absorption band shapes between 3.2 and 3.6 μm that are similar to those of Themis, Ceres, and other main-belt asteroids (Fig. 1B).

We also produced a globally averaged thermal infrared spectrum using data from the

Fig. 1. Comparison of 3.4- μm absorption features on main-belt asteroids and multiple locations on Bennu. (A) C-complex asteroid spectra obtained at 15° to 30° phase angle (22)—except the (24) Themis data from (18)—each plotted with 1σ uncertainties. Curves are offset from 1.0 for clarity. (B) The OVIRS globally averaged Bennu spectrum at 9° phase angle shows a broad absorption feature of $\sim 2\%$ depth, extending from 3.2 to 3.6 μm . The solid line shows the spectrum smoothed with a nine spectral channel-width Gaussian. Spectra of individual locations on Bennu are also shown, the coordinates of each location labeled, and curves are offset from 1.0 for clarity. Uncertainties are smaller than the symbol size.



OSIRIS-REx Thermal Emission Spectrometer (OTES) (23), acquired during the Detailed Survey phase, to search for spectral features that could be attributable to carbon in organics and/or carbonates. However, other minerals exhibit features in the same spectral region, complicating our interpretation. The radiometric performance of OTES is degraded at $>1300\text{ cm}^{-1}$ ($\lesssim 7.7\text{ }\mu\text{m}$) and at surface temperatures lower than 325 K (23), requiring us to co-add spectra to maintain an adequate signal-to-noise ratio (fig. S1) (24). This prevents us from obtaining detailed spatial distribution information for comparison with the OVIRS observations made at four times the areal resolution. We therefore limit our analyses to just the OVIRS data.

Feature repeatability

With repeated coverage over some areas of Bennu's surface, we investigated the effects of phase angle on spectral shape to determine whether the absorption is an observational or processing artifact. The globally averaged spectra from the three local solar times are shown in Fig. 2A. A similar $3.4\text{-}\mu\text{m}$ band shape is seen in each average spectrum, although at 12:30 p.m. and 3:00 p.m. the feature is slightly deeper near $3.45\text{ }\mu\text{m}$ and has more absorption below $3.3\text{ }\mu\text{m}$ than at 10:00 a.m. There is no apparent thermal emission fill-in in the absorption band at any phase angle, because the average surface temperatures were low (328 K at 12:30 p.m.); however, individual hot surface locations become very noisy in the 3:00 p.m. spectra. Using the average surface temperatures, reflectance, and solar range-corrected flux for each observation, the radiance contribution from reflected solar light at $3.4\text{ }\mu\text{m}$ is shown to be ~ 31 , 22, and 15% for 10:00 a.m., 12:30 p.m., and 3:00 p.m., respectively. However, the absolute reflected solar radiance at 12:30 p.m. is nearly twice that at 10:00 a.m. and 3:00 p.m., resulting in the highest signal-to-noise ratio after thermal tail subtraction (fig. S2) (24).

We also investigated feature repeatability on a regional scale. Using the regions with higher-resolution observations (Fig. 2B), we averaged several overlapping spots together to obtain a close (although not identical) match to the spatial resolution of each Detailed Survey map. Two overlapping spots in each observation cover $\sim 900\text{ m}^2$, and the afternoon spectra again have a deeper band than the 10:00 a.m. spectra, but spectral shapes are similar for each observation. We also compared the average of all high-spatial-resolution spots over an area with the corresponding low-spatial-resolution spectrum (Fig. 2, C and D). Although the spectra were taken at different local solar times, differences between the resulting spectra are $<0.5\%$, indicating that phase angle does not appreciably affect the band depth or shape.

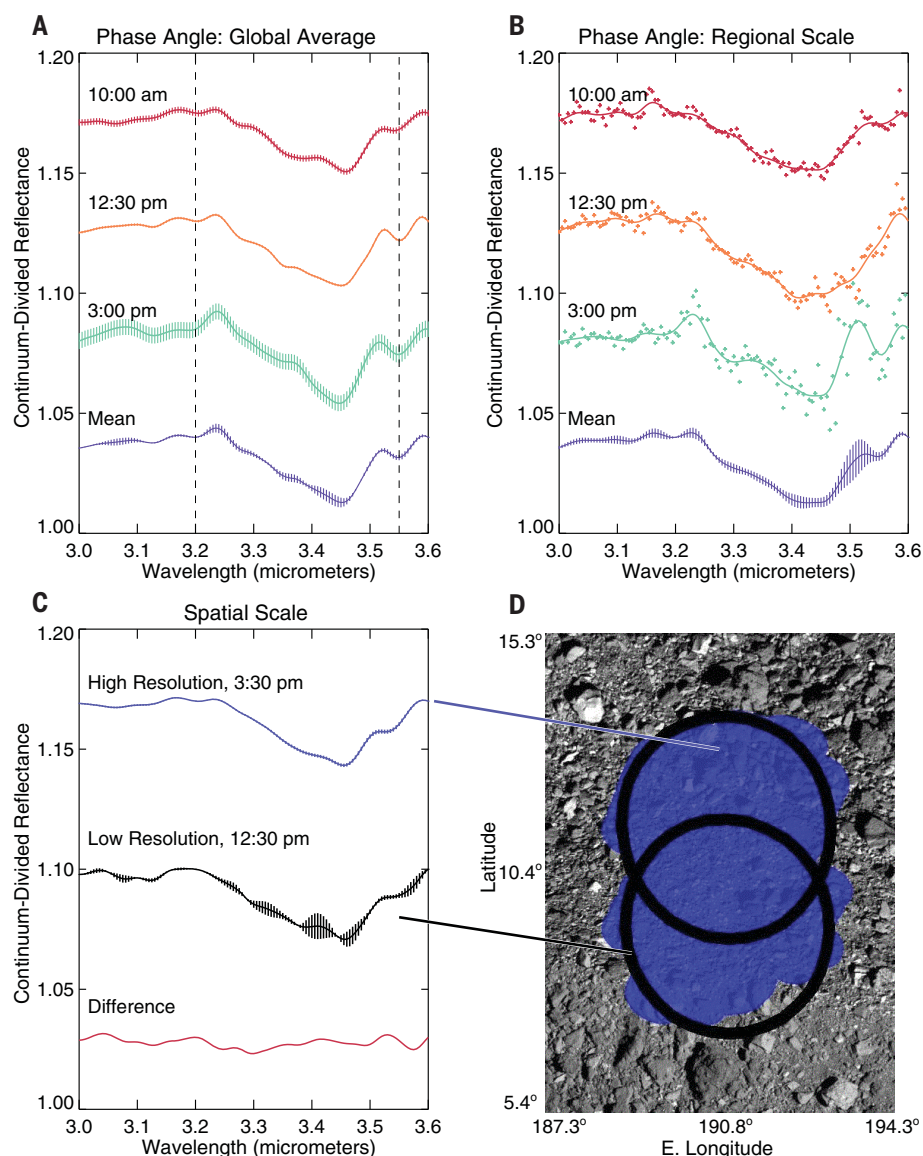


Fig. 2. Effect of local solar time (phase angle) and spatial scale on absorption feature shapes and depths. (A) Globally averaged spectra at 10:00 a.m. ($\sim 30^\circ$ phase), 12:30 p.m. ($\sim 9^\circ$), and 3:00 p.m. (45°) local solar time, and the mean across all three phase angles; hatching indicates the standard error of the mean. (B) Regional-scale ($\sim 900\text{ m}^2$) spectra at the same phase angles as the global spectra, and their mean with standard error (hatching). Points show the data, and lines are after smoothing with a nine spectral channel-width Gaussian. (C) Average of 163 local spectra (blue) compared with a regional two-spectrum average (black) and their difference (red); spectra are plotted as Gaussian-smoothed curves (lines) with standard error (hatching). (D) Approximate areas covered by the observations shown in (C), but without vertical along-track smear, which improves the overlap between the regional (black empty circles) and local (blue filled circles) spectral spots.

Global surface distribution

To map the distribution of the absorption feature across the surface, we used the 12:30 p.m. spectra, as they have the highest signal-to-noise ratio and fewest shadows. From those data, we calculated the band area (band depth integrated from 3.2 to $3.6\text{ }\mu\text{m}$), mapped it onto the surface of a three-dimensional shape model of Bennu (25), and overlaid the results on an OSIRIS-REx Camera Suite (OCAMS) base map with 5-cm spatial resolution (26, 27) (Fig. 3A).

Band area was chosen over single-channel band depth to cover the range of spectral shapes that could arise as a result of different carbonates or organics (28).

The map in Fig. 3A shows that carbon-bearing material is spread over 98% of Bennu's surface, with band areas varying from 0.12 to 0.82% , but in no discernible large-scale pattern; 98% of the surface between $\pm 50^\circ$ latitude shows a band depth of $>1\%$ at $3.42\text{ }\mu\text{m}$, and 32% of the surface has an absorption $\geq 2\%$. Most other

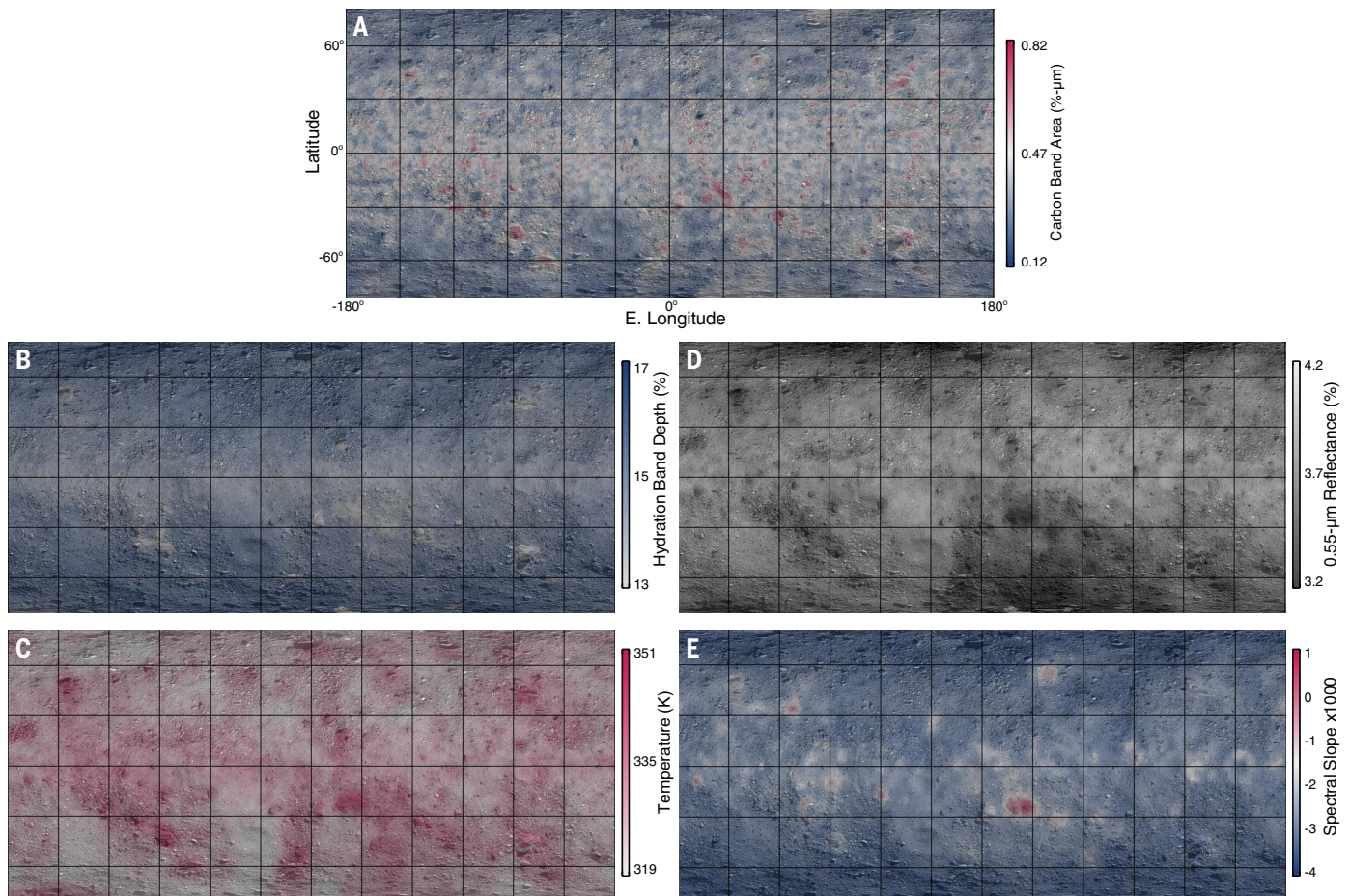


Fig. 3. Global spectral maps of Bennu from the OVIRS observations taken at the 12:30 p.m. Equatorial Station. (A) Band area from 3.2 to 3.6 μm , showing absorption due to carbon-bearing materials. (B) 2.74- μm hydration band depth, (C) effective surface temperature, (D) 0.55- μm reflectance, and (E) spectral slope from 0.5 to 1.5 μm . The latitude and longitude grid shown in (A) is overlaid on the other panels.

surface parameters, such as 12:30 p.m. surface temperature, 0.55- μm reflectance, or visible spectral slope (Fig. 3, B to D), align with surface features: the darkest boulders and blanketed areas (lowest 0.55- μm reflectance) tend to be the hottest and have the shallowest hydration band at 2.74 μm . The boulders appear to have low thermal inertia, high thermal roughness, and usually a redder spectral slope (29, 30).

Over wavelengths from 0.5 to 1.5 μm , most of the surface has a blue (negative) spectral slope (8, 9), but some individual regions have a redder (flat to slightly positive) slope (Fig. 3E). Some individual boulders stand out in the spectral slope map, but not all boulders are redder than average; some are bluer (steeper negative slope). These data agree with photometric color observations at higher spatial resolution, which show that individual boulders and craters have a variety of colors (30). Many of the areas that are redder than average occur at low latitudes, in agreement with some ground-based measurements that find equatorial reddening, although the measured

variation in slope in OVIRS data is much smaller than is seen in those studies (31). As with the 3.4- μm absorption, the range of spectral behavior observed on Bennu at wavelengths shorter than 2.5 μm is similar to that found among B-type asteroids throughout the main belt (5). Bennu's blue color may in part be due to space weathering (30).

To determine whether Bennu's global characteristics are correlated at low spatial resolution, we calculated Spearman's rank correlation values, r , on different map pairs. We limited the calculation to latitudes less than 50° to avoid effects from variable solar illumination on short wavelength and thermal parameters, which affect high latitudes more strongly; at this limit, the temperature map has weak correlation to incidence or emission angles ($|r| < 0.25$), and further narrowing the latitude range did not meaningfully affect the results. At low phase angle, maps of the 2.74- μm hydration band depth and 0.55- μm reflectance show some correlation with peak surface temperature ($|r| = 0.76$ and 0.65 , respectively) (fig. S3) (24). The slope from 0.5 to 1.5 μm shows the

least correlation with peak temperature, with the bluest and the reddest regions both being warmer than the average. However, the spectral slope does weakly correlate with the hydration band and 0.55- μm reflectance ($|r| > 0.55$), with the reddest large boulders having the shallowest hydration band and the lowest reflectance.

Although the 3.2- to 3.6- μm band area map shows that some boulders have a deeper absorption than others, comparison with other surface parameters (24) indicates no obvious correlation, and all map pairs have $|r| \leq 0.45$. Bennu's carbon band area follows neither the reddest nor the darkest regions on the surface. This would be unexpected if Bennu's 3.4- μm band were attributable entirely to organics (32). However, as most spectra have an absorption shape that matches a mix of organics and carbonates (28), we cannot globally map these independently of one another. Thus, the lack of correlation may be the result of composition or for another reason, such as particle size effects.

As the 0.55- μm albedo and hydration band depth are correlated with surface temperature, we also produced a detrended 2.74- μm

hydration band map (fig. S4) to search for residual variations after removing the temperature dependence (24). Neither the band area from 3.2 to 3.6 μm nor the spectral slope from 0.5 to 1.5 μm show any correlation with the detrended hydration map, with $|r| \leq 0.23$ (24). The 2.74- μm hydration band map, caused by the presence of hydrated phyllosilicates, shows a correlation with latitude, $|r| = 0.58$ (fig. S5). This correlation decreases to $|r| = 0.48$ in the temperature-detrended map. The correlations between hydration band and peak surface temperature or latitude could be due to dehydration, differential space weathering, compositional differences, particle size, surface roughness, or other causes. Although hydrated phyllosilicates and carbon-bearing material are globally present, their lack of correlation indicates separate formation processes.

Local variability

At the 20- to 30-m spot scale used in the global maps, spot-to-spot variability in band area (0.04 ± 0.01 to $0.84 \pm 0.02\%$ μm) indicates a heterogeneous surface, either in terms of band depth or shape. At the 5- to 10-m scale, the band shape and area are consistent over small areas where the spots overlap on a geologic feature (Fig. 4B) but change as the spots move to adjacent rocks, possibly because of local variability (Fig. 4, A and C). However, some of these high-spatial-resolution data were acquired at phase angles where shadows may affect the signal, particularly at high latitudes or near large boulders (Fig. 4A).

Nonetheless, at the finer spatial scale, absorption depths exceed 10% in a few isolated locations; these locations with the deepest absorptions likely have the highest concentra-

tion of carbon-bearing material, although viewing conditions and particle size effects may also affect band depth. While full global coverage was not obtained at this resolution, 187 spectra (~1% of the high-spatial-resolution spots) were found to have absorptions of 5 to 14% at 3.42, 3.45, or 3.47 μm ; these wavelengths were chosen to span the deepest absorption wavelengths for a mix of organics and carbonates (28). Most of these locations occur over boulders; only a single spectrum with a deep absorption feature occurs over a smooth crater floor, which is ringed by boulders having deep absorptions. However, this region is at high latitude, and the viewing conditions included some large shadows.

To avoid potential effects on band depth caused by shadows and low signal, we identified 18 low-latitude spectra with a 3.4- μm absorption

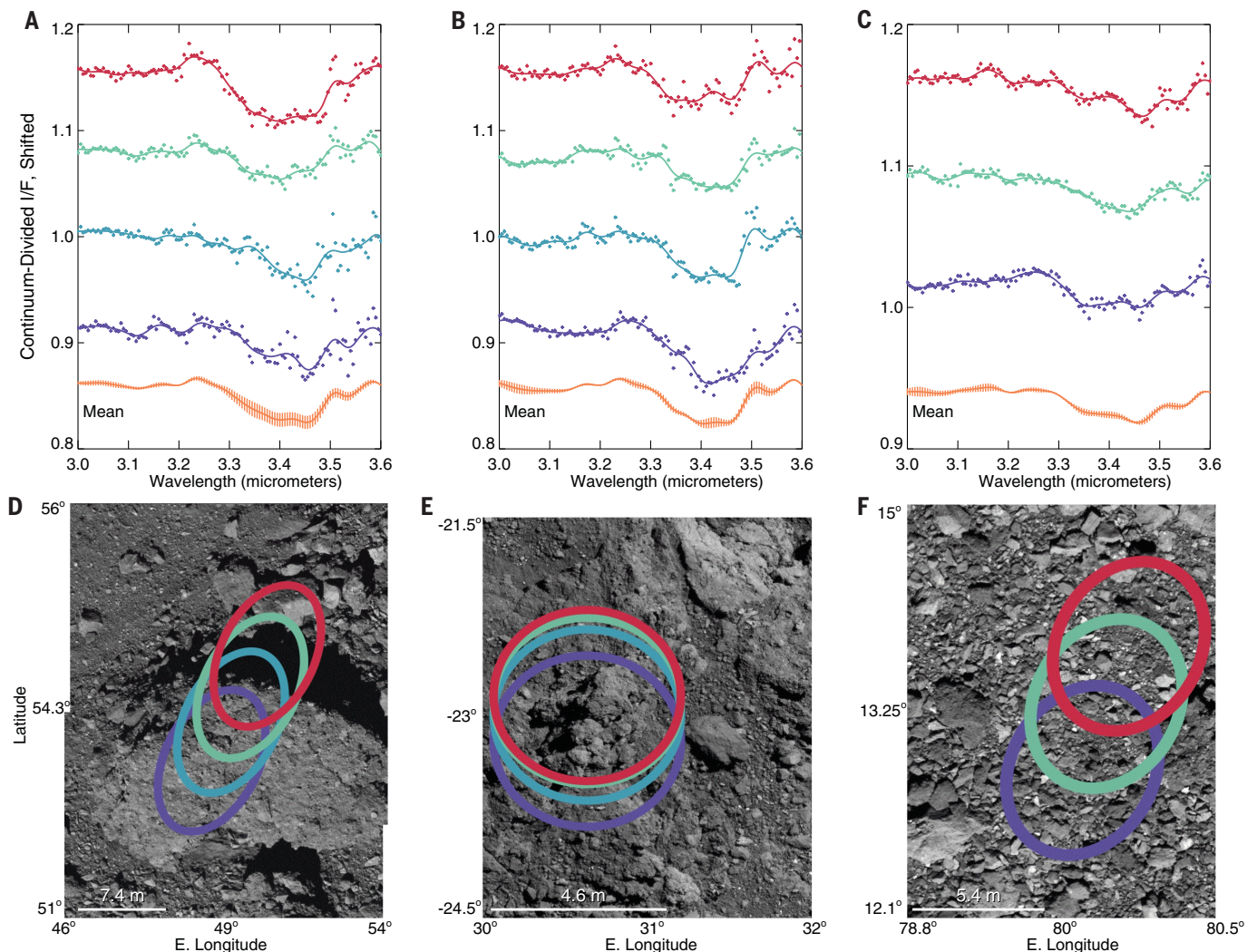


Fig. 4. Individual high-areal-resolution (~60 m²) local spectra and their corresponding OVIRS footprints. (A to C) Spectra of selected locations and their means. (D to F) The corresponding OVIRS footprints (ellipses, elongated by along-track motion) overlaid on OCAMS panchromatic images acquired at the same time. In (A), a deep absorption feature changes spectral shape and depth as the field of view moves off a large boulder and onto smaller boulders (D), which could have different composition, and the mean shows variation over the full band. In (B), a cluster of spectra over a rock pile (E) shows consistent shapes with less variation in the mean than in (A). In (C), a region with little absorption shows few large boulders in the field of view (F) and similar variation in the mean as (B).

feature >5% in depth and overlaid them on the OCAMS map of Bennu (26, 27). In some instances, the spots overlap spatially, resulting in 10 individual locations with deep 3.4- μm absorption (Fig. 5A). All of these locations lie on large boulders that dominate the OVIRS field of view. As with the global-scale maps, these boulders span a range of brightness and colors, with no distinguishing morphological features; some appear buried or are composites of broken material, whereas others are neither (30). Another low-latitude region (Fig. 5B) shows that the deepest absorptions occur in a rock field between two craters. In comparison, the OSIRIS-REx mission's primary sample site, Nightingale crater at 56°N, 42°E (Fig. 5C), shows many spots with 3.4- μm absorption of 5 to 7% depth, despite low solar illumination. Again, although there is an ab-

sorption over the entire region, most spots with deep features correspond to boulders around the rim of the crater (28).

Origin of carbon material

The variation in 3.4- μm feature shape depends on surface location, with individual spot spectra that match those of other C-complex asteroids (Fig. 1A), demonstrating a heterogeneity that spans multiple asteroid families and classes. This heterogeneity is likely caused by varying proportions of carbonate and organics (28), which we cannot distinguish in our global dataset. A spectral feature we observe at 3.98 μm may be due to carbonates (28) (Fig. 6), but the signal-to-noise ratio of the thermal-corrected spectra is <25 above 3.7 μm (24). An observed 2.3- μm feature could also be due to carbonates or to phyllosilicates, but the lack of a corre-

sponding 2.5- μm absorption makes an identification uncertain. Spectral slopes could be affected by the presence of organics (33), but Bennu's spectral slope variation is small, and redder OVIRS spectra show similar absorption features to bluer spectra in the low-resolution 600 m^2 data (Fig. 6).

We do not observe widespread evidence of a 3.1- μm absorption from water ice or ammoniated phyllosilicates (Fig. 1B), which is present on many C-complex main-belt asteroids (e.g., Ceres and Themis) (18, 22), indicating that Bennu's bulk mineralogy and chemistry are distinct, despite the similarities in the 3.4- μm region. As a rubble-pile near-Earth asteroid, Bennu is a mixture of parent-body material aggregates that have been warmed by proximity to the Sun, which would have removed volatile water ice (34).

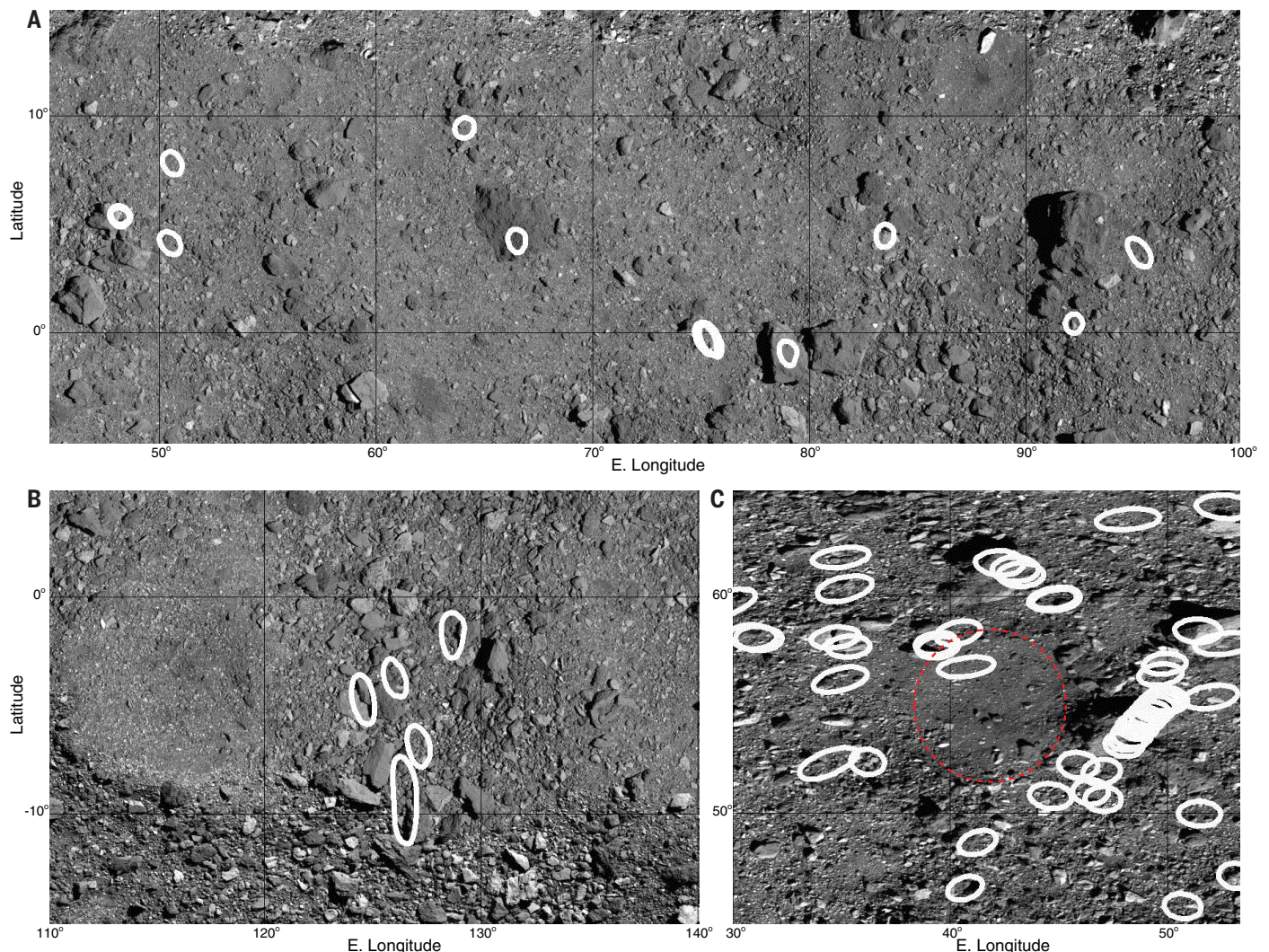
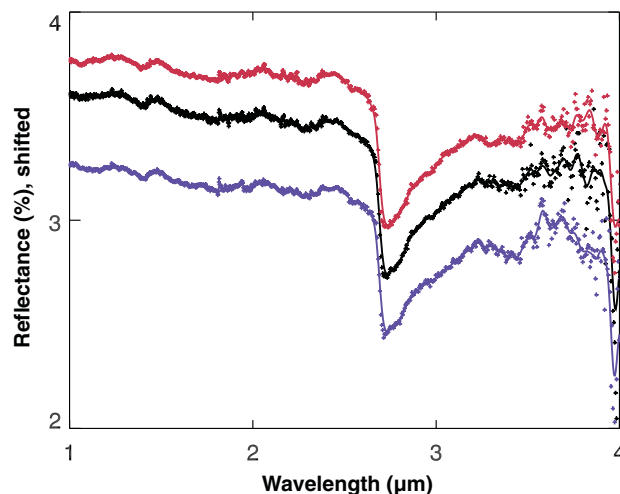


Fig. 5. The locations and sizes of OVIRS footprints (white outlines) of spectra with single-channel absorptions >5% at either 3.42, 3.45, or 3.47 μm . The footprints are overlaid on the 5-cm-resolution base map (26, 27); they are elongated by surface projection effects but neglect along-track smear. Footprint coverage over the images is sparse, primarily in separated north-south scans. (A) Low latitudes show a number of deep absorptions, primarily over boulders. (B) An adjacent equatorial region shows that the boulders between two craters also have deep absorptions. (C) The environs around the OSIRIS-REx primary sample site, Nightingale crater (red dashed circle), also show locally deeper absorptions.

Fig. 6. Bennu reflectance spectra from 2 to 4 μm .

The global average spectrum (black curve) shows other absorption features centered near 2.3 and 3.98 μm . There is little difference in the reflectance spectrum of any spot, regardless of spectral slope from 0.5 to 1.5 μm , with redder regions (red curve) showing the same spectral features as bluer regions (blue curve). The red and blue curves are shifted by 0.2 for clarity.



Local regions on Ceres have complex mineralogy including organic matter, carbonates, and ammoniated phases (33, 35). Ground-based observations of Ceres indicated the presence of carbonates, while orbital data showed several organic-rich areas (32, 36). The limited spatial concentration of aliphatic organics on Ceres has been attributed to geologically recent exposure of those materials, such that they have not been broken down by solar irradiation (33, 35, 37). With irradiation, aliphatic carbon breaks down to graphitized or amorphous carbon, and these phases are thought to be ubiquitous on Ceres (38); they may also be present on Bennu (29) but cannot be directly observed with NIR spectroscopy. The widespread presence of aliphatic organics at 3.4 μm would indicate that neither space weathering nor heating on Bennu were pervasive enough to degrade (or graphitize) all aliphatic organics. To preserve a widespread 3.4- μm organic feature, most of the material on Bennu's surface could not have been exposed to the space environment for more than a few million years (39–41).

Two scenarios could explain the carbon-bearing material on Bennu's surface and its variety of spectral shapes: accretion of external material over time or gradual exposure of a mix of materials present from its aggregation. First, given the similarity of local-area spectral shapes to different C-complex asteroids, Bennu's surface may exhibit a mix of materials accreted during its time in the main belt and later transit to a near-Earth orbit (exogenous material). Pyroxene-bearing boulders have been detected on the surface and are spectrally consistent with the howardite, eucrite, and diogenite achondrite meteorite groups associated with asteroid (4) Vesta (42), indicating that the surface retains at least some exogenous material. However, at the scale of these observations (~20 m per spot), there is no obvious connection between the carbon-bearing absorption features and any particu-

lar rock type (30). The meteorites that are the closest spectral analogs to Bennu (CI and CM chondrites) tend to have less than 5 vol % carbonates. Carbon-bearing materials were likely ubiquitous in the forming solar system and interstellar medium, so they would have been available to form organics and carbonates in planetesimals (43). Any such material delivered to the surface of Bennu would be altered by space weathering and high surface temperatures, especially during perihelion, so pristine carbon-bearing material, particularly organics, would not be preserved unless synthesized and delivered within the past million years (39–41). The widespread distribution of such material on Bennu's surface makes this scenario unlikely.

Alternatively, as a rubble pile formed after a collision (2, 25), Bennu's surface may be peppered with boulders aggregated from different layers of its parent body and/or the impactor. Again, any surface materials, especially organics, would degrade over time. It is possible that buried organic material is being newly exposed by surface impacts, although we do not observe any correlation between the carbon-bearing material and craters (Figs. 3 and 5).

In the carbon-bearing boulders, thermal cycling and degradation by meteoroid impacts could cause fatigue and fracturing that exposes new organic material over time (44, 45). The ejection of particles from Bennu's surface supports the idea that fresh surfaces are actively being exposed (46). Bennu's blue spectral slope is also consistent with the colors of a subset of asteroids that show evidence of ongoing mass loss and are known as main-belt comets. However, on Bennu, this blue spectral slope extends to the shortest wavelengths we observe, which is rare for main-belt comets (19). Conversely, some of Bennu's boulders display veins that may be composed of carbonate minerals (28), lending further credence to the supposition that we are observing a mixture of carbon-bearing materials.

Our observations provide context for the sample planned to be returned to Earth by OSIRIS-REx. The widespread surface distribution of the 3.4- μm absorption on Bennu indicates that the sample should contain carbon-bearing material, regardless of sampling location. The primary sample site, Nightingale, shows deeper regional carbon-bearing absorptions than average, particularly in the boulders around the crater rim; this does not rule out substantial carbon-bearing material on the crater floor.

Methods summary

All OVIRS spectra were calibrated using an automated pipeline that converts raw counts to calibrated radiance units (watts per square centimeter per micrometer per steradian) (24, 47). Before analysis, each spectrum was resampled onto a common wavelength axis with 2-nm spectral resolution below 2.4 μm and 5-nm resolution from 2.4 to 4.3 μm . These resampled spectra were then fitted with a model of a thermal tail, which was subtracted, then divided by the solar flux to produce reflectance spectra (24). Surface feature maps were constructed using a Lommel-Seeliger disk correction to remove latitudinal effects. For band depth and shape comparisons, the background continuum was removed using a model fitted to each spectrum (24).

REFERENCES AND NOTES

1. D. N. DellaGiustina *et al.*, Properties of rubble-pile asteroid (101955) Bennu from OSIRIS-REx imaging and thermal analysis. *Nat. Astron.* **3**, 341–351 (2019). doi: [10.1038/s41550-019-0731-1](https://doi.org/10.1038/s41550-019-0731-1)
2. D. S. Lauretta *et al.*, The OSIRIS-REx target asteroid (101955) Bennu: Constraints on its physical, geological, and dynamical nature from astronomical observations. *Meteorit. Planet. Sci.* **50**, 834–849 (2015). doi: [10.1111/maps.12353](https://doi.org/10.1111/maps.12353)
3. D. J. Tholen, "Asteroid taxonomy from cluster analysis of photometry," thesis, University of Arizona, Tucson (1984).
4. S. J. Bus, P. R. Binzel, Phase II of the Small Main-Belt Asteroid Spectroscopic Survey, A feature-based taxonomy. *Icarus* **158**, 146–177 (2002). doi: [10.1006/icar.2002.6856](https://doi.org/10.1006/icar.2002.6856)
5. J. de León, N. Pinilla-Alonso, H. Campins, J. Licandro, G. A. Marzo, Near-infrared spectroscopic survey of B-type asteroids: Compositional analysis. *Icarus* **218**, 196–206 (2012). doi: [10.1016/j.icarus.2011.11.024](https://doi.org/10.1016/j.icarus.2011.11.024)
6. D. S. Lauretta *et al.*, OSIRIS-REx: Sample return from asteroid (101955) Bennu. *Space Sci. Rev.* **212**, 925–984 (2017). doi: [10.1007/s11214-017-0405-1](https://doi.org/10.1007/s11214-017-0405-1)
7. B. E. Clark *et al.*, Spectroscopy of B-type asteroids: Subgroups and meteorite analogs. *J. Geophys. Res.* **115**, E06005 (2010). doi: [10.1029/2009JE003478](https://doi.org/10.1029/2009JE003478)
8. B. E. Clark *et al.*, Asteroid (101955) 1999 RQ36: Spectroscopy from 0.4 to 2.4 μm and meteorite analogs. *Icarus* **216**, 462–475 (2011). doi: [10.1016/j.icarus.2011.08.021](https://doi.org/10.1016/j.icarus.2011.08.021)
9. V. E. Hamilton *et al.*, Evidence for widespread hydrated minerals on asteroid (101955) Bennu. *Nat. Astron.* **3**, 332–340 (2019). doi: [10.1038/s41550-019-0722-2](https://doi.org/10.1038/s41550-019-0722-2); pmid: [31360777](https://pubmed.ncbi.nlm.nih.gov/31360777/)
10. W. F. Bottke *et al.*, In search of the source of asteroid (101955) Bennu: Applications of the stochastic YORP model. *Icarus* **247**, 191–217 (2015). doi: [10.1016/j.icarus.2014.09.046](https://doi.org/10.1016/j.icarus.2014.09.046)
11. H. Campins *et al.*, The origin of asteroid 101955 (1999 RQ36). *Astrophys. J.* **721**, L53–L57 (2010). doi: [10.1088/2041-8205/721/L53](https://doi.org/10.1088/2041-8205/721/L53)
12. N. Pinilla-Alonso *et al.*, Portrait of the Polana-Eulalia family complex: Surface homogeneity revealed from near-infrared spectroscopy. *Icarus* **274**, 231–248 (2016). doi: [10.1016/j.icarus.2016.03.022](https://doi.org/10.1016/j.icarus.2016.03.022)
13. S. Sugita *et al.*, The geomorphology, color, and thermal properties of Ryugu: Implications for parent-body processes.

- Science* **364**, eaaw0422 (2019). doi: [10.1126/science.aaw0422](https://doi.org/10.1126/science.aaw0422); pmid: [30890587](https://pubmed.ncbi.nlm.nih.gov/30890587/)
14. K. Kitazato *et al.*, The surface composition of asteroid 162173 Ryugu from Hayabusa2 near-infrared spectroscopy. *Science* **364**, 272–275 (2019). doi: [10.1126/science.aaw7432](https://doi.org/10.1126/science.aaw7432); pmid: [30890589](https://pubmed.ncbi.nlm.nih.gov/30890589/)
 15. H. Campins *et al.*, Spectra of asteroid families in support of Gaia. *Planet. Space Sci.* **73**, 95–97 (2012). doi: [10.1016/j.pss.2012.06.017](https://doi.org/10.1016/j.pss.2012.06.017)
 16. D. C. Reuter *et al.*, The OSIRIS-REx Visible and InfraRed Spectrometer (OVIRS): spectral maps of the asteroid Bennu. *Space Sci. Rev.* **214**, 54 (2018). doi: [10.1007/s11214-018-0482-9](https://doi.org/10.1007/s11214-018-0482-9)
 17. B. Hapke, "Thermal emission and emission spectroscopy" in *Theory of Reflectance and Emittance Spectroscopy* (Cambridge Univ. Press, ed. 2, 2012), pp. 412–439.
 18. A. S. Rivkin, J. P. Emery, Detection of ice and organics on an asteroidal surface. *Nature* **464**, 1322–1323 (2010). doi: [10.1038/nature09028](https://doi.org/10.1038/nature09028); pmid: [20428165](https://pubmed.ncbi.nlm.nih.gov/20428165/)
 19. H. Campins *et al.*, Water ice and organics on the surface of the asteroid 24 Themis. *Nature* **464**, 1320–1321 (2010). doi: [10.1038/nature09029](https://doi.org/10.1038/nature09029); pmid: [20428164](https://pubmed.ncbi.nlm.nih.gov/20428164/)
 20. A. S. Rivkin, E. S. Howell, J. P. Emery, Infrared spectroscopy of large, low-albedo asteroids: Are Ceres and Themis archetypes or outliers? *J. Geophys. Res. Planets* **124**, 1393–1409 (2019). doi: [10.1029/2018JE005833](https://doi.org/10.1029/2018JE005833)
 21. A. S. Rivkin, E. L. Volquardsen, B. E. Clark, The surface composition of Ceres: Discovery of carbonates and iron-rich clays. *Icarus* **185**, 563–567 (2006). doi: [10.1016/j.icarus.2006.08.022](https://doi.org/10.1016/j.icarus.2006.08.022)
 22. F. Usui, S. Hasegawa, T. Ootsubo, T. Onaka, AKARI/IRC near-infrared asteroid spectroscopic survey: AcuA-spec. *Publ. Astron. Soc. Jpn.* **71**, 1–41 (2019). doi: [10.1093/pasj/psy125](https://doi.org/10.1093/pasj/psy125)
 23. P. R. Christensen *et al.*, The OSIRIS-REx Thermal Emission Spectrometer (OTES) instrument. *Space Sci. Rev.* **214**, 87 (2018). doi: [10.1007/s11214-018-0513-6](https://doi.org/10.1007/s11214-018-0513-6)
 24. Materials, methods, and related figures are available as supplementary materials.
 25. O. S. Barnouin *et al.*, Shape of (101955) Bennu indicative of a rubble pile with internal stiffness. *Nat. Geosci.* **12**, 247–252 (2019). doi: [10.1038/s41561-019-0330-x](https://doi.org/10.1038/s41561-019-0330-x); pmid: [31080497](https://pubmed.ncbi.nlm.nih.gov/31080497/)
 26. B. Rizk *et al.*, OCAMS: The OSIRIS-REx Camera Suite. *Space Sci. Rev.* **214**, 26 (2018). doi: [10.1007/s11214-017-0460-7](https://doi.org/10.1007/s11214-017-0460-7)
 27. C. A. Bennett *et al.*, A high-resolution global basemap of (101955) Bennu. *Icarus* **113690** (2020). doi: [10.1016/j.icarus.2020.113690](https://doi.org/10.1016/j.icarus.2020.113690)
 28. H. H. Kaplan *et al.*, Bright carbonate veins on asteroid (101955) Bennu: Implications for aqueous alteration history. *Science* **370**, eabc3557 (2020). doi: [10.1126/science.abc3557](https://doi.org/10.1126/science.abc3557); pmid: [33033155](https://pubmed.ncbi.nlm.nih.gov/33033155/)
 29. B. Rozitis *et al.*, Asteroid (101955) Bennu's weak boulders and thermally anomalous equator. *Sci. Adv.* **6**, eabc3699 (2020). doi: [10.1126/sciadv.abc3699](https://doi.org/10.1126/sciadv.abc3699); pmid: [33033037](https://pubmed.ncbi.nlm.nih.gov/33033037/)
 30. D. N. DellaGiustina *et al.*, Variations in color and reflectance on the surface of asteroid (101955) Bennu. *Science* **370**, eabc3660 (2020). doi: [10.1126/science.abc3660](https://doi.org/10.1126/science.abc3660); pmid: [33033157](https://pubmed.ncbi.nlm.nih.gov/33033157/)
 31. R. P. Binzel *et al.*, Spectral slope variations for OSIRIS-REx target Asteroid (101955) Bennu: Possible evidence for a fine-grained regolith equatorial ridge. *Icarus* **256**, 22–29 (2015). doi: [10.1016/j.icarus.2015.04.011](https://doi.org/10.1016/j.icarus.2015.04.011)
 32. C. M. Pieters *et al.*, Geologic constraints on the origin of red organic-rich material on Ceres. *Meteorit. Planet. Sci.* **53**, 1983–1998 (2018). doi: [10.1111/maps.13008](https://doi.org/10.1111/maps.13008)
 33. H. H. Kaplan, R. E. Milliken, C. M. O'D. Alexander, New constraints on the abundance and composition of organic matter on Ceres. *Geophys. Res. Lett.* **45**, 5274–5282 (2018). doi: [10.1029/2018GL077913](https://doi.org/10.1029/2018GL077913)
 34. B. Rozitis *et al.*, Implications for ice stability and particle ejection from high-resolution temperature modeling of asteroid (101955) Bennu. *J. Geophys. Res. Planets* **125**, (2020). doi: [10.1029/2019JE006323](https://doi.org/10.1029/2019JE006323)
 35. M. C. De Sanctis *et al.*, Characteristics of organic matter on Ceres from VIR/Dawn high spatial resolution spectra. *Mon. Not. R. Astron. Soc.* **482**, 2407–2421 (2019). doi: [10.1093/mnras/sty2772](https://doi.org/10.1093/mnras/sty2772)
 36. M. C. De Sanctis *et al.*, Localized aliphatic organic material on the surface of Ceres. *Science* **355**, 719–722 (2017). doi: [10.1126/science.aaj2305](https://doi.org/10.1126/science.aaj2305); pmid: [28209893](https://pubmed.ncbi.nlm.nih.gov/28209893/)
 37. S. Marchi *et al.*, An aqueously altered carbon-rich Ceres. *Nat. Astron.* **3**, 140–145 (2019). doi: [10.1038/s41550-018-0656-0](https://doi.org/10.1038/s41550-018-0656-0)
 38. A. R. Hendrix, F. Vilas, J.-Y. Li, Ceres: Sulfur deposits and graphitized carbon. *Geophys. Res. Lett.* **43**, 8920–8927 (2016). doi: [10.1002/2016GL070240](https://doi.org/10.1002/2016GL070240)
 39. B. E. Clark, B. Hapke, C. Pieters, D. Britt, "Asteroid space weathering and regolith evolution" in *Asteroids III*, W. F. Bottke Jr., A. Cellino, P. Paolicchi, R. P. Binzel, Eds. (Univ. of Arizona Press, 2002), pp. 585–599.
 40. R. Brunetto, M. J. Loeffler, D. Nesvorný, S. Sasaki, G. Strazzulla, "Asteroid surface alteration by space weathering processes" in *Asteroids IV*, P. Michel, F. E. DeMeo, W. F. Bottke, Eds. (Univ. of Arizona Press, 2015), pp. 597–616.
 41. C. M. Pieters, S. K. Noble, Space weathering on airless bodies. *J. Geophys. Res. Planets* **121**, 1865–1884 (2016). doi: [10.1002/2016JE005128](https://doi.org/10.1002/2016JE005128); pmid: [29862145](https://pubmed.ncbi.nlm.nih.gov/29862145/)
 42. D. N. DellaGiustina *et al.*, Exogenic basalt on asteroid (101955) Bennu. *Nat. Astron.* **10.1038/s41550-020-1195-z** (2020). doi: [10.1038/s41550-020-1195-z](https://doi.org/10.1038/s41550-020-1195-z)
 43. F. J. Ciesla, S. A. Sandford, Organic synthesis via irradiation and warming of ice grains in the solar nebula. *Science* **336**, 452–454 (2012). doi: [10.1126/science.1217291](https://doi.org/10.1126/science.1217291); pmid: [22461502](https://pubmed.ncbi.nlm.nih.gov/22461502/)
 44. J. L. Molaro, S. Byrne, J.-L. Le, Thermally induced stresses in boulders on airless body surfaces, and implications for rock breakdown. *Icarus* **294**, 247–261 (2017). doi: [10.1016/j.icarus.2017.03.008](https://doi.org/10.1016/j.icarus.2017.03.008)
 45. J. L. Molaro *et al.*, In situ evidence of thermally induced rock breakdown widespread on Bennu's surface. *Nat. Commun.* **11**, 2913 (2020). doi: [10.1038/s41467-020-16528-7](https://doi.org/10.1038/s41467-020-16528-7); pmid: [32518333](https://pubmed.ncbi.nlm.nih.gov/32518333/)
 46. D. S. Lauretta *et al.*, Episodes of particle ejection from the surface of the active asteroid (101955) Bennu. *Science* **366**, eaay3544 (2019). doi: [10.1126/science.aay3544](https://doi.org/10.1126/science.aay3544); pmid: [31806784](https://pubmed.ncbi.nlm.nih.gov/31806784/)
 47. A. A. Simon *et al.*, In-flight calibration and performance of the OSIRIS-REx Visible and IR Spectrometer (OVIRS). *Remote Sens.* **10**, 1486 (2018). doi: [10.3390/rs10091486](https://doi.org/10.3390/rs10091486)
 48. P. R. Christensen, E. Engle, S. Anwar, S. Dickenshied, D. Noss, N. Gorelick, M. Weiss-Malik, "JMARS – A Planetary GIS," American Geophysical Union Fall Meeting, San Francisco, CA, 14 to 18 December 2009; <https://ui.adsabs.harvard.edu/abs/2009AGUFMIN22A..06C/abstract>.
 49. D. C. Reuter, A. A. Simon, A. Lunsford, D. S. Lauretta, Origins, Spectral Interpretation, Resource Identification, Security, Regolith Explorer (OSIRIS-REx): Visible and InfraRed Spectrometer (OVIRS) Bundle. NASA Planetary Data System (2019); <https://sbn.psi.edu/pds/resource/orex/ovirs.html>.

ACKNOWLEDGMENTS

We are grateful to the entire OSIRIS-REx Team for making the encounter with Bennu possible and to C. W. V. Wolner for editorial review. We thank the developers of the JMARS open source software (https://jmars.mars.asu.edu/open_source) for assistance with Bennu-specific visualization (48). **Funding:** A.A.S., H.H.K., V.E.H., D.S.L., H.C., J.P.E., D.C.R., D.N.D., S.A.S., D.R.G., L.F.L., A.R., and C.A.B. were supported by NASA under contract NNM10AA11C issued through the New Frontiers Program. B.R. acknowledges funding support from the Royal Astronomical Society (RAS) and the UK Science and Technology Facilities Council (STFC). **Author contributions:** Investigation: A.A.S., D.S.L., D.C.R., V.E.H., and H.H.K. Data validation: A.A.S., D.C.R., V.E.H., and D.R.G. Data curation: A.A.S., D.C.R., H.H.K., V.E.H., D.R.G., D.N.D., and C.A.B. Conceptualization: A.A.S., D.S.L., D.C.R., V.E.H., H.H.K., J.P.E., A.R., B.R., M.A.B., L.F.L., and S.A.S. Methodology: A.A.S., D.S.L., D.C.R., V.E.H., H.H.K., J.P.E., B.R., M.A.B., L.F.L., and S.A.S. Formal analysis: A.A.S., H.H.K., and V.E.H. Writing or editing: A.A.S., H.H.K., V.E.H., D.S.L., H.C., J.P.E., D.C.R., D.N.D., S.A.S., D.R.G., L.F.L., A.R., B.R., M.A.B., and C.A.B. **Competing interests:** We declare no competing interests. **Data and materials availability:** All OVIRS spectral data from the Detailed Survey and Reconnaissance phases are available via the Planetary Data System at <https://sbn.psi.edu/pds/resource/orex/ovirs.html> (49). Full resolution maps and corresponding 1 σ uncertainty maps are available in FITS format as data S1 to S6.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6517/eabc3522/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S6
References (50–52)
Data S1 to S6

20 April 2020; accepted 27 August 2020
Published online 8 October 2020
[10.1126/science.abc3522](https://doi.org/10.1126/science.abc3522)

RESEARCH ARTICLE SUMMARY

ASTEROIDS

Bright carbonate veins on asteroid (101955) Bennu: Implications for aqueous alteration history

H. H. Kaplan*, D. S. Lauretta, A. A. Simon, V. E. Hamilton, D. N. DellaGiustina, D. R. Golish, D. C. Reuter, C. A. Bennett, K. N. Burke, H. Campins, H. C. Connolly Jr., J. P. Dworkin, J. P. Emery, D. P. Glavin, T. D. Glotch, R. Hanna, K. Ishimaru, E. R. Jawin, T. J. McCoy, N. Porter, S. A. Sandford, S. Ferrone, B. E. Clark, J.-Y. Li, X.-D. Zou, M. G. Daly, O. S. Barnouin, J. A. Seabrook, H. L. Enos

INTRODUCTION: Carbonaceous asteroids formed early in Solar System history and experienced varying degrees of aqueous (water-rock) and thermal alteration. Most models of the evolution of these asteroids suggest that aqueous alteration was driven by hydrothermal convection. However, it is debated whether this alteration occurred in a chemically closed or open system. The bulk chemical compositions of the carbonaceous chondrite meteorites imply that the system was closed. Models predict that large-scale fluid flow in an open system took place on at least some asteroids. In this scenario, fluids would have flowed through fractures from the interior, and minerals would have precipitated into these fractures, forming veins.

RATIONALE: Global spectral observations by the OSIRIS-REx (Origins, Spectral Interpretation, Resource Identification, and Security-

Regolith Explorer) spacecraft have shown that carbon-bearing materials, including organics and/or carbonates, are widespread on the surface of near-Earth asteroid (101955) Bennu. To understand the composition of these carbon-bearing materials and their implications for Bennu's aqueous alteration history, we examined visible-near-infrared spectra obtained at resolutions of ~4 m in the region around the mission's primary sampling site, Nightingale. We compared these spectra with laboratory spectra of organics and carbonates. Broad-band panchromatic images of the Nightingale region at pixel scales of ~1.4 cm allowed us to search for surface features that could contextualize the spectral data.

RESULTS: On the basis of the wavelength position and shape of spectral absorption features, the closest matches to a subset (~15%) of

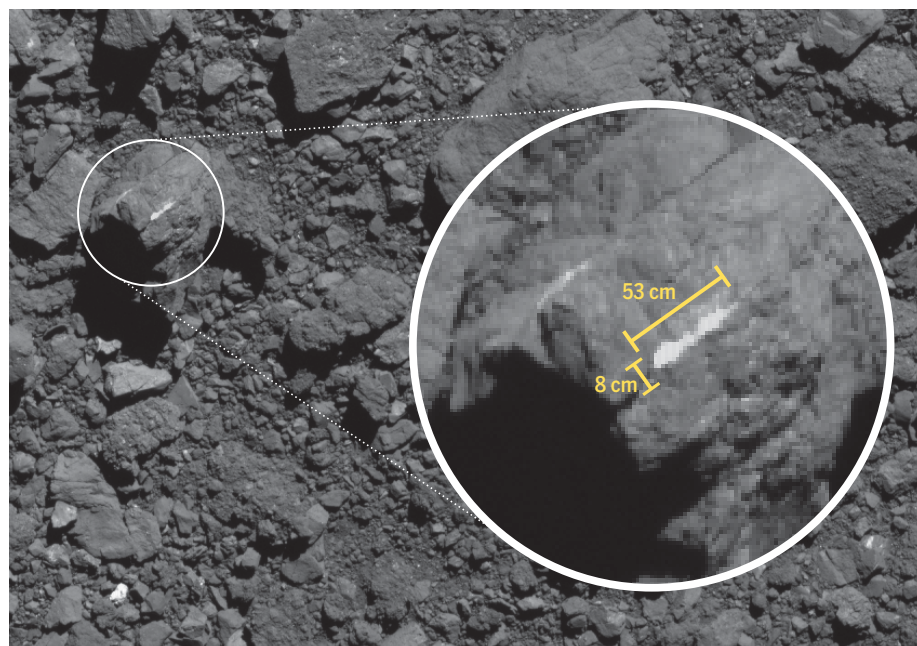
spectra collected in the Nightingale region are the carbonate minerals calcite, dolomite, and magnesite, with calcite being the most prevalent. These minerals are found in the highly aqueously altered CM- and CI-group carbonaceous chondrites, proposed meteorite analogs of Bennu.

Bright, centimeters-thick, roughly meter-long veins are apparent in images of some boulders near the Nightingale site. The veins have albedos of 10% to at least 19%, much brighter than Bennu's average of 4.4%. These albedos are consistent with the veins being filled by carbonates with minor abundances of opaque materials such as magnetite and organics. The albedos are too high for the veins to be dominantly composed of organics or any other material that has been identified on Bennu, except carbonates.

There are only a few bright vein-filling materials found in meteorites, and carbonate is the only one of these that is plausibly present on Bennu. Linear mixture modeling suggests that only a small fraction of a spectrometer field of view (<1%) needs to contain carbonate to explain the observed spectral signatures, which is consistent with the observed extent of the veins.

CONCLUSION: The observed bright veins on Bennu are most likely dominantly composed of carbonates. The detection of these meter-sized, putative carbonate veins indicates that extensive water flow occurred in a chemically open hydrothermal system on the larger parent asteroid of Bennu, before the catastrophic disruption of the parent asteroid led to Bennu's reaccumulation as a rubble pile. Kinetic modeling assuming a calcite vein composition shows that a hydrothermal system kilometers in size, with alteration likely taking place over thousands to millions of years, would have been needed to create veins of the dimensions preserved in the constituent boulders of Bennu.

Some CM- and CI-group meteorites have very small (micrometers to millimeters) carbonate veins or presumed vein fragments; veins on the scale that we observe on Bennu are not present in meteorites. It is possible that such veins are present on other carbonaceous asteroids that have not been observed by spacecraft. We predict that the sample of Bennu's surface material that is planned to be returned to Earth by the OSIRIS-REx spacecraft will contain carbonates and that they will be distributed with a structure and scale distinct from those that occur in meteorites. ■



OSIRIS-REx panchromatic image (~1.4 cm pixel⁻¹) showing a boulder with a bright vein on Bennu.

This linear feature is more than twice as bright as the surrounding rock. Its most plausible composition is carbonate, suggesting that a large, long-lived hydrothermal system operated on Bennu's parent asteroid. The image was collected by the PolyCam imager on 26 October 2019 at 61.60°N, 50.57°E.

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

*Corresponding author. Email: hannah.kaplan@nasa.gov

Cite this article as H. H. Kaplan *et al.*, *Science* **370**, eabc3557 (2020). DOI: 10.1126/science.abc3557

S READ THE FULL ARTICLE AT
<https://doi.org/10.1126/science.abc3557>

RESEARCH ARTICLE

ASTEROIDS

Bright carbonate veins on asteroid (101955) Bennu: Implications for aqueous alteration history

H. H. Kaplan^{1,2*}, D. S. Lauretta³, A. A. Simon¹, V. E. Hamilton², D. N. DellaGiustina³, D. R. Golish³, D. C. Reuter¹, C. A. Bennett³, K. N. Burke³, H. Campins⁴, H. C. Connolly Jr.^{5,3}, J. P. Dworkin¹, J. P. Emery⁶, D. P. Glavin¹, T. D. Glotch⁷, R. Hanna⁸, K. Ishimaru³, E. R. Jawin⁹, T. J. McCoy⁹, N. Porter³, S. A. Sandford¹⁰, S. Ferrone¹¹, B. E. Clark¹¹, J.-Y. Li¹², X.-D. Zou¹², M. G. Daly¹³, O. S. Barnouin¹⁴, J. A. Seabrook¹³, H. L. Enos³

The composition of asteroids and their connection to meteorites provide insight into geologic processes that occurred in the early Solar System. We present spectra of the Nightingale crater region on near-Earth asteroid Bennu with a distinct infrared absorption around 3.4 micrometers. Corresponding images of boulders show centimeters-thick, roughly meter-long bright veins. We interpret the veins as being composed of carbonates, similar to those found in aqueously altered carbonaceous chondrite meteorites. If the veins on Bennu are carbonates, fluid flow and hydrothermal deposition on Bennu's parent body would have occurred on kilometer scales for thousands to millions of years. This suggests large-scale, open-system hydrothermal alteration of carbonaceous asteroids in the early Solar System.

The Origins, Spectral Interpretation, Resource Identification, and Security–Regolith Explorer (OSIRIS-REx) mission to asteroid (101955) Bennu is designed to return a carbon-rich sample of the asteroid to Earth (1). Bennu, a 500-m-diameter near-Earth asteroid, was chosen as the mission target because of its spectral similarity to primitive and organic-rich carbonaceous chondrite meteorites (2). Primitive chondritic meteorites formed in the early Solar System, so they may record the materials, processes, and events that occurred during that period.

The OSIRIS-REx spacecraft entered orbit around Bennu in December 2018. Throughout 2019 and early 2020, the spacecraft acquired images and spectra to characterize Bennu's surface properties, composition, and relationship to meteorites, in advance of sample collection.

Bennu is spectrally classified as a B-type asteroid on the basis of its blue (negative) spectral slope. B-types are a subset of the

larger C-complex of primitive asteroids, so named for their presumably carbonaceous compositions. Bennu's surface contains abundant, widespread hydrated minerals (3), has widespread magnetite (3, 4), and is dominated by boulders up to ~100 m in longest dimension (5). Near-infrared spectra of Bennu's surface exhibit an absorption feature due to hydroxyl in hydrated clay minerals; the band minimum is 2.74 μm , most consistent with hydrated Mg-bearing phyllosilicates (3, 6). Thermal infrared spectra exhibit silicate stretching and bending modes consistent with volumetrically dominant phyllosilicates, as well as absorptions attributable to magnetite near 18 and 29.4 μm (555 and 340 cm^{-1}) (3). These observations demonstrate that Bennu is mineralogically similar to the carbonaceous chondrite meteorites that show evidence of water-rock interaction (i.e., aqueous alteration), with the closest meteorite analogs being the most heavily altered members of the CM and CI groups. Aqueous alteration of these meteorites likely took place on a large parent asteroid, >30 km in diameter (7), early in Solar System history (8). Dynamical analyses suggest that Bennu, a rubble pile asteroid, consists of fragments of a parent body of ~100-km diameter that reaccumulated after a catastrophic disruption (9–11).

Establishing Bennu's relationship with specific groups of meteorites—whether CI, CM, or others—would provide insight into the geologic processes that Bennu's parent body experienced. The CI and CM groups record diverse scenarios of parent body aqueous alteration. All known CI chondrites are categorized on the basis of their mineral and rock properties as petrologic type 1 (CI1) (12). This classifies the

CI1s as heavily hydrated, and they contain the highest abundances of volatile elements of any meteorites (13). CM chondrites span a broader range of degrees of aqueous alteration, from petrologic type 1 to 2 (14, 15). The CM chondrites are depleted in volatiles relative to the CI1s. Although the CM1 and CI1 meteorites are both highly aqueously altered, they have distinct bulk chemistry and textures (15). There are other carbonaceous chondrites that cannot be classified into any of the known groups and are therefore termed “ungrouped” (16). Grouping chondrites together implies that they likely originate from the same geologic context; if a carbonaceous chondrite is ungrouped, it is possible that it has a distinct geologic origin. We consider recent observations of asteroid Bennu within this meteorite framework, to constrain its geologic history and make predictions for the sample that will be returned to Earth.

Spectra of carbon species on Bennu

In global observations of Bennu collected during the Detailed Survey–Equatorial Stations campaign of the OSIRIS-REx mission (25 April to 6 June 2019) (1), the OSIRIS-REx Visible and InfraRed Spectrometer (OVIRS) (17) detected a ubiquitous infrared absorption feature near 3.4 μm (18), a spectral region associated with carbon-bearing species. OVIRS measures reflected light at wavelengths from 0.4 to 4.3 μm with a circular, 4-mrad field of view (17). Between 3.2 and 3.6 μm , there are at least five absorption minima attributed to organic molecules, including symmetric and asymmetric stretching modes of methyl ($-\text{CH}_3$) and methylene ($-\text{CH}_2$) groups (i.e., aliphatic CH), and an aromatic CH stretch. Carbonates have overlapping absorption features in this wavelength region from overtones (absorptions occurring at higher-energy multiples of the fundamental frequency) and combinations of the fundamental stretches in CO_3 .

The global data have a spatial resolution of ~30 m in the direction of spacecraft motion (along-track) and 20 m in the perpendicular direction (cross-track), with coverage of nearly the entire asteroid surface (18). The complex spectral features near 3.4 μm in this dataset are attributed at the global scale to mixtures of organics and carbonates (18). The spectral shape in this wavelength region is repeatable to <1% across multiple observations of the same surface spot, indicating that instrumental and illumination effects do not cause the observed absorption complexity (18).

We analyzed higher-spatial-resolution OVIRS observations that were collected in October 2019 during the first Reconnaissance (Recon A) phase of the OSIRIS-REx mission, which entailed more detailed investigations of candidate sampling sites. We used data from the region around Nightingale crater, the mission's

¹NASA Goddard Space Flight Center, Greenbelt, MD, USA.

²Southwest Research Institute, Boulder, CO, USA. ³Lunar and Planetary Laboratory, University of Arizona, Tucson, AZ, USA.

⁴Department of Physics, University of Central Florida, Orlando, FL, USA.

⁵Department of Geology, School of Earth and Environment, Rowan University, Glassboro, NJ, USA.

⁶Department of Astronomy and Planetary Sciences, Northern Arizona University, Flagstaff, AZ, USA.

⁷Department of Geosciences, Stony Brook University, Stony Brook, NY, USA.

⁸Jackson School of Geosciences, University of Texas, Austin, TX, USA.

⁹Smithsonian Institution National Museum of Natural History, Washington, DC, USA.

¹⁰NASA Ames Research Center, Mountain View, CA, USA.

¹¹Department of Physics and Astronomy, Ithaca College, Ithaca, NY, USA.

¹²Planetary Science Institute, Tucson, AZ, USA.

¹³Centre for Research in Earth and Space Science, York University, Toronto, Ontario, Canada.

¹⁴John Hopkins University Applied Physics Laboratory, Laurel, MD, USA.

*Corresponding author. Email: hannah.kaplan@nasa.gov

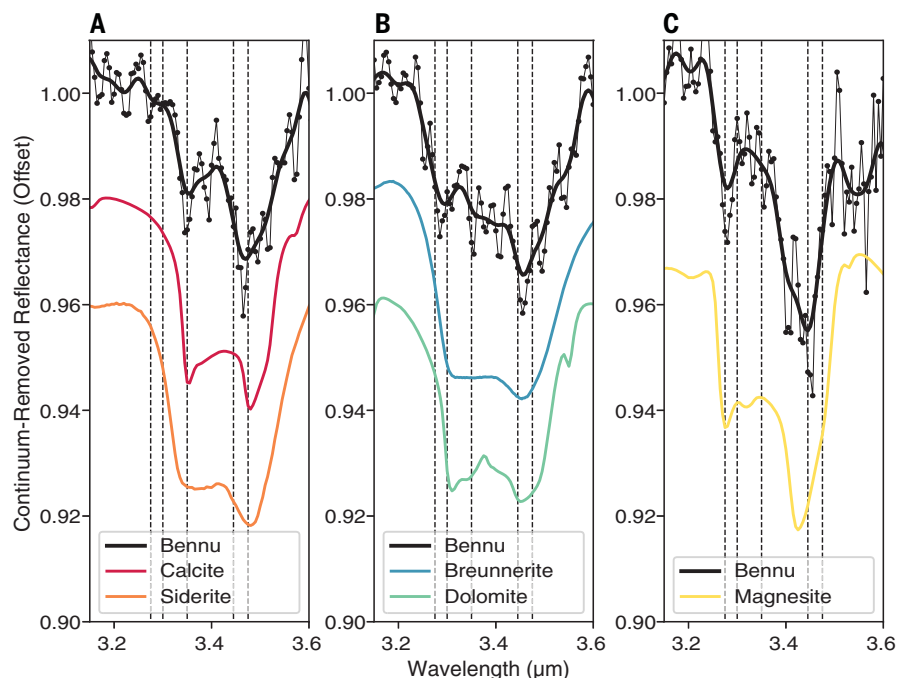


Fig. 1. Spectral features in OVIRS data compared to laboratory carbonate spectra. (A to C) Black points show the original OVIRS spectrum of Bennu from three different locations in Nightingale, and the thick black lines are the spectrum smoothed with a 3σ Gaussian. The carbonate laboratory spectra (colored lines as indicated in the legends) are natural, terrestrial samples with variable cation content (table S1). All spectra have had their continuum removed and are offset.

primary sampling site, to provide context for the sample return. These observations have spatial resolutions of ~ 4 m cross-track and 9 m along-track for each OVIRS footprint.

The shapes of the $3.4\text{-}\mu\text{m}$ feature in the Nightingale observations differ from those in the global observations (18), indicating that it may be possible to distinguish compositions at higher spatial resolution. To discriminate between organics and carbonates, we use a chi-square goodness of fit test, band minima positions, and visual inspection (19). The goodness of fit test uses a set of 220 laboratory spectra of carbonates and organics with a range of compositions to assess how closely these laboratory data match the shape of the OVIRS data (table S1). Uncertainties were determined by adding noise to the laboratory spectra to mimic the OVIRS signal-to-noise ratio (SNR) near $3.4\text{ }\mu\text{m}$ (19). We find that chi-square values of <1 can identify either carbonate minerals or organics with $>99\%$ confidence ($>95\%$ confidence for chi-square <2) (table S1).

Organic and carbonate spectral features differ in terms of number of minima, minima positions, and absorption widths. We identify some OVIRS spectra with $3.4\text{-}\mu\text{m}$ features that are consistent with carbonates and others that are consistent with organics (Figs. 1 and 2). The OVIRS spectral shapes consistent with organics are similar to those previously detected near $3.4\text{ }\mu\text{m}$ in asteroids [e.g., (20–22)] and me-

teorites (23) (Fig. 2). These features have a narrow minimum at $3.42\text{ }\mu\text{m}$ or a broad, flat feature centered at $3.42\text{ }\mu\text{m}$ [e.g., (24)]. The OVIRS spectral signature consistent with carbonates is primarily identified by an absorption feature with two minima near $3.4\text{ }\mu\text{m}$ (Fig. 1). The band positions vary by tens of nanometers, consistent with known changes in wavelength of carbonate absorptions with varying cation composition (Fig. 3). The OVIRS spectra that are best fitted with carbonate laboratory spectra (Fig. 2) are poorly fitted with organic laboratory spectra (fig. S1).

The carbonate-matching OVIRS spectra include features similar to those of calcite (CaCO_3), siderite (FeCO_3), magnesite (MgCO_3), dolomite [$\text{CaMg}(\text{CO}_3)_2$], and/or breunnerite [$(\text{Mg,Fe,Mn})\text{CO}_3$] (Figs. 1 and 2). We cannot distinguish calcite and siderite at the spectral resolution and SNR of the OVIRS data near $3.4\text{ }\mu\text{m}$; however, Fe in siderite would produce a broad absorption near $1\text{ }\mu\text{m}$, which is not detected (fig. S2). Siderite is rare in carbonaceous chondrite meteorites, whereas calcite is abundant (24). We therefore conclude that calcite is the most likely carrier responsible for these carbonate-matching spectra. Aragonite has the same composition as calcite (CaCO_3) with a different crystal structure; its spectral properties are similar to those of calcite at wavelengths $<2.5\text{ }\mu\text{m}$ (25), but appropriate laboratory spectra are not available, so it is

not included in our analysis. Spectra that we identify as dolomite could alternatively be due to Fe-rich breunnerite, which has an absorption feature at a similar wavelength. For more Mg-rich forms of breunnerite, we expect the band position to shift toward shorter wavelengths, similar to magnesite. We have not found a $3.4\text{-}\mu\text{m}$ feature on Bennu that corresponds to natrite (Na_2CO_3), which was detected in the brightest regions of dwarf planet Ceres and has minima at longer wavelengths than calcite (26).

In addition to the feature at $3.4\text{ }\mu\text{m}$, carbonates are expected to produce an absorption near $4\text{ }\mu\text{m}$. OVIRS has an SNR of 50 at $3.4\text{ }\mu\text{m}$, but the signal decreases rapidly beyond $\sim 3.6\text{ }\mu\text{m}$, with an SNR of only 15 at $4\text{ }\mu\text{m}$. Low SNR reduces our ability to detect the expected carbonate absorption near $4\text{ }\mu\text{m}$, but we do find a possible absorption feature in this region that is correlated with the $3.4\text{-}\mu\text{m}$ feature (fig. S2). Overtones at ~ 2.35 and, possibly, $2.55\text{ }\mu\text{m}$ are also observed in the OVIRS spectrum (fig. S2) and may be due to carbonate and/or Mg or Fe phyllosilicates. However, these features have $<2\%$ band depth, the strength of the absorption feature compared to the continuum, and any variation in band depth is within the uncertainty.

Thermal infrared spectra from the OSIRIS-REx Thermal Emission Spectrometer (OTES) (27) exhibit a spectral feature near $7.04\text{ }\mu\text{m}$ (1420 cm^{-1}) that may be attributable to carbonates (28), but this feature has not been definitively assigned (3). Another carbonate feature, the ν_3 (third fundamental) absorption, is expected at ~ 6.35 to $6.75\text{ }\mu\text{m}$ (1500 cm^{-1}) and has been observed in laboratory and remote-sensing data (28, 29); however, it has not been identified in OTES global spectral observations (spatial resolutions of 20 to 40 m). Nor have other expected carbonate features, including the narrow, ν_2 fundamental absorption near $11.3\text{ }\mu\text{m}$ (886 cm^{-1}) and a large, broad, fundamental absorption between ~ 25 and $33\text{ }\mu\text{m}$ (~ 400 to 300 cm^{-1}). Detection of spectral features in the thermal infrared depends on SNR, temperature, physical conditions (e.g., roughness or cavities), phase abundance, the compositions of the other phases that are present, spectral resolution, and particle size (19). These factors may explain the nondetection of these carbonate bands, despite the higher SNR of the OTES data relative to that of the OVIRS data.

We mapped a $\sim 0.03\text{-km}^2$ area surrounding the Nightingale sample site ($\sim 56.04^\circ$ latitude, 42.05° longitude), within which $\sim 1\%$ of the 3396 OVIRS spectra have a chi-square <1 match to laboratory spectra of carbonates, rather than organics or mixtures, and $\sim 15\%$ have chi-square <2 (Fig. 2) (19). Organics account for $<1\%$ (chi-square <1) or 8% (chi-square <2) of the spectra based on these goodness of fit tests. The spectral signatures with chi-square <1 match

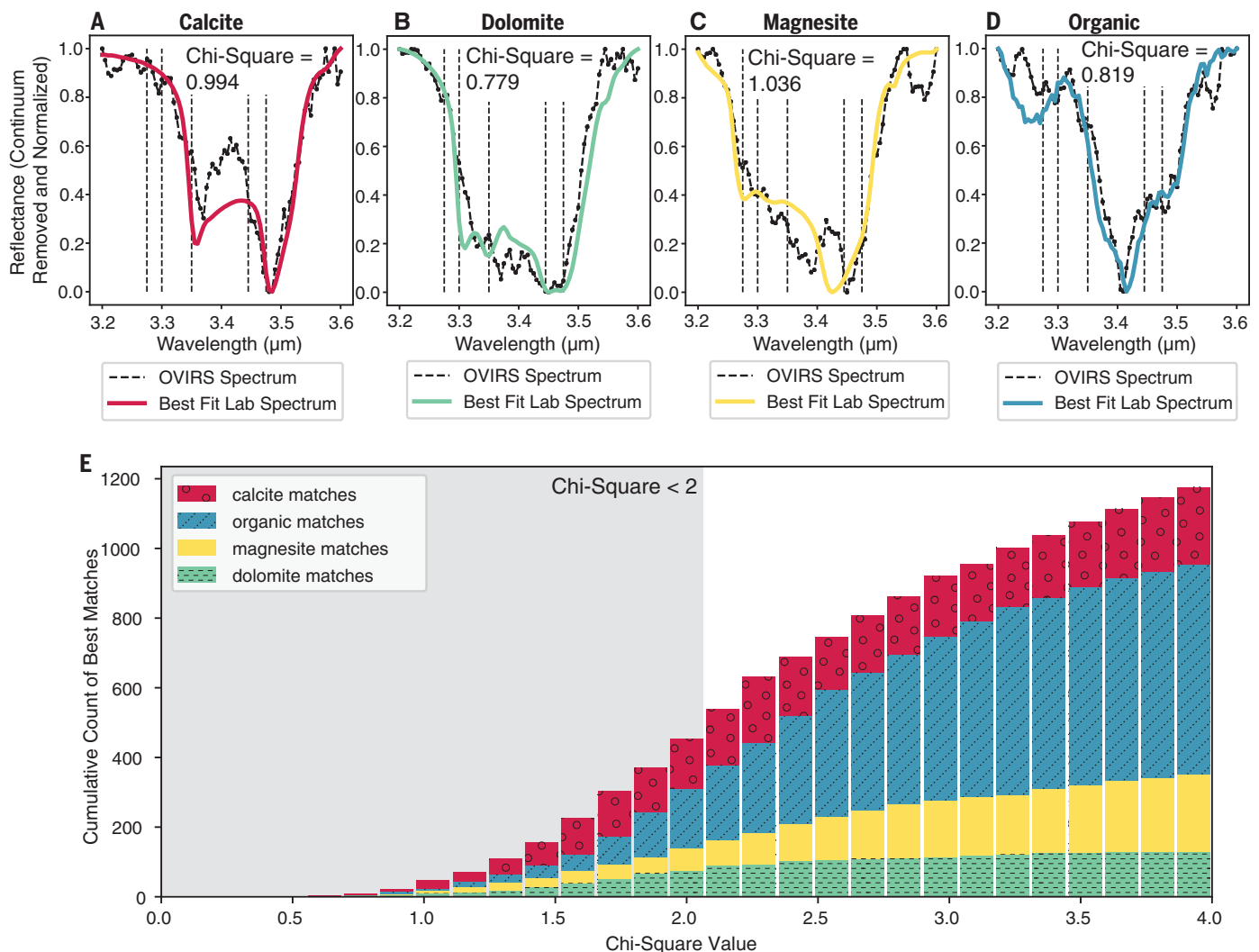


Fig. 2. Goodness of fit tests for laboratory carbonate spectra compared to OVIRS spectra. (A to D) Four different OVIRS spectra in the 3.4- μm region fit that are best fitted with laboratory spectra of calcite, dolomite, magnesite, and meteorite insoluble organic matter, respectively. (E) The cumulative count of all OVIRS spectra best fitted with either a calcite, dolomite, magnesite, or organic laboratory spectrum ordered by the chi-square value of the model. We discuss matches with chi-square < 2 (gray-shaded region) in the text.

to carbonates are strongest at the rim of Nightingale crater and on individual boulders in the area. Some of these same surface features are also associated with the strongest absorptions at 3.2 to 3.6 μm , though many are not (Fig. 4). The majority (~64%) of the carbonate signatures have band positions that are consistent with calcite, followed by magnesite/Mg-rich breunnerite (~24%) and dolomite/Fe-rich breunnerite (12%) (Figs. 2 and 3).

Bright vein-like features and mottling in boulders

To provide geologic context for the spectral observations, we examined high-resolution (~1.4 cm pixel⁻¹) images of the Nightingale region acquired by the OSIRIS-REx Camera Suite (OCAMS) PolyCam imager (30, 31) during Recon A. We found centimeter- to meter-scale features on boulders that are distinctly brighter

than the surrounding host rock. Some of these features are elongated and linear, whereas others are irregularly shaped. They are apparent in the raw camera images, with digital number (DN) values between 1.4 and 1.9 times the average value of the host rock (table S3), indicating higher reflectance than the average Benu surface.

To quantify the brightness and morphology of these features, we focused on three host boulders (Fig. 5). We radiometrically calibrated the images using the mission's standard calibration process (31). This process converts raw DN values to reflectance (also known as radiance factor or I/F), the ratio of incident solar radiation that is reflected from the asteroid to that which would be reflected from an ideal Lambertian surface normally illuminated by the Sun. The reflectance values are then used to determine the bright material's

normal albedo, defined as the reflectance at normal observing conditions (0° solar incidence, 0° solar phase, and 0° emission angles). We registered the calibrated images to digital terrain models (DTMs) of the boulders, previously produced from OSIRIS-REx Laser Altimeter (OLA) data with an average facet size of ~7 cm (19, 32). We estimated the observed photometric angles by ray tracing from the registered position and orientation of the camera to the DTMs. We used the derived angles as inputs to the Robotic Lunar Observatory photometric correction (33) to compensate for the reflectance variations resulting from changing viewing and illumination conditions. We then measured the absolute difference in normal albedo between the bright features and their host rocks.

The median normal albedo of Benu, ~4.4%, is among the lowest of any object in the Solar

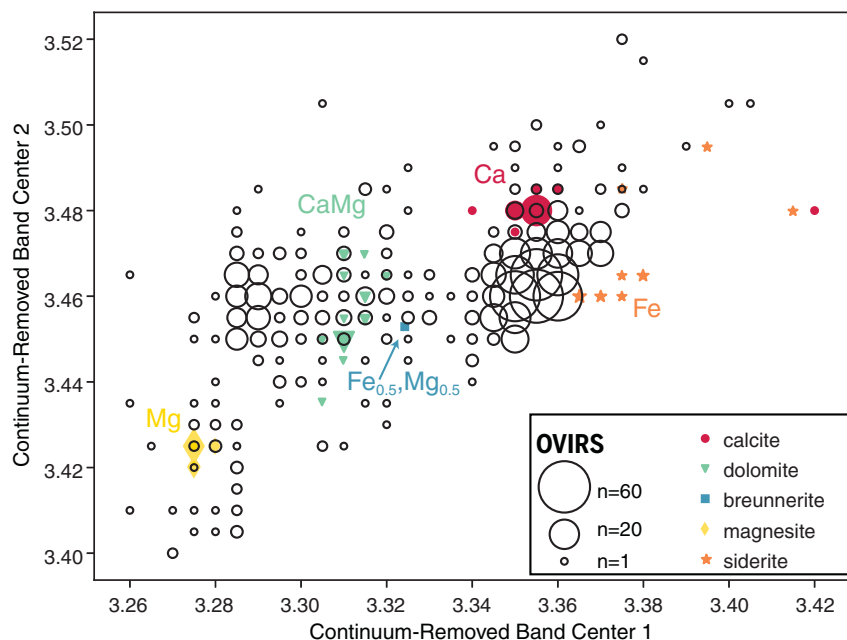


Fig. 3. Band minima positions of OVIRS signatures compared to laboratory spectra of carbonates. The positions of two carbonate band minima near 3.4 μm are plotted for Bennu (empty black circles), along with values for laboratory spectra (colored solid symbols; see legend and table S1). The OVIRS minima positions indicate a range of compositions on the surface. Circles are scaled to show the number of spectra represented, as indicated in the legend.

System (5, 34). Analyses of the photometrically corrected images show that the bright features have normal albedos of at least 10 to 19% (Fig. 5 and table S3). The mean normal albedo of the host boulders varies between 6.0 and 7.6% (table S3), which is also brighter than Bennu's global average.

We performed further analysis of one of the bright feature-bearing boulders (labeled VBR-13 in Fig. 5, where VBR stands for vein-bearing rock) to assess the uncertainty and variability of the derived normal albedo values. This boulder was identified in 13 images taken at varying resolutions and observation geometries (table S3). In images with pixel scales ranging from 1.3 to 4.8 cm pixel^{-1} , the normal albedo of the bright material ranges from 14 to 19% ($\pm 2.2\%$, 1σ) (table S3). We rule out textural and observation geometry effects and conclude that the bright features are high-reflectance material included in the host rocks.

The three illustrative examples of the elongated, linear bright features are 3 to 15 cm thick and vary in length from 50 to 150 cm, though smaller examples exist (Fig. 6). These occur at sharp edges of boulders so may be small exposures of a larger internal layer. We interpret all these features as veins—narrow, linear features formed when an aqueous solution flows through a rock, depositing minerals by precipitation, usually within preexisting fractures. Shock processes can also create linear or planar features, and melt veins created by shock

events are found in some meteorites (35). Shock melt veins are darker than the surrounding meteorite matrix, whereas the veins on Bennu are brighter than their host rocks.

The irregularly shaped bright features in darker host rocks produce a mottled appearance (Fig. 6). Similar mottled boulders occur on another C-complex asteroid, (162173) Ryugu (36). The irregular features on Bennu are typically >10 cm in the longest dimension, range from angular to rounded, and can occur individually or in clusters.

The host rocks of the veins and irregular features span a range of albedos and sizes but tend to have similar morphologies. The host boulders appear to be part of a previously identified population that has a distinct, brighter albedo distribution (though not as bright as the veins and irregular features they host) relative to the dark boulders that dominate Bennu's surface, and a more negative spectral slope at ultraviolet (UV) wavelengths (37). These brighter boulders also have a different texture, appearing smoother and more angular than the dark boulders on Bennu. The former exhibit fractures and pits and appear to be harder and less easily broken down than the other, more hummocky observed boulder morphologies. The largest bright boulder is ~ 10 to 12 m; all of the largest boulders (≥ 20 m) on the asteroid are dark (5). In many cases, boulders that host bright veins or irregular features do so only on a single face. A few small boulders have bright material that dominates an entire face (fig. S3).

Analysis of their colors is published elsewhere (37). They have higher thermal inertias than the dark boulders, which may indicate lower porosity (38), as we expect for precipitation of vein-filling materials in fractures and void spaces.

Carbonates and veins in meteorites

Although meteorites are a biased sample of the asteroids from which they are derived, comparison may offer insight into the possible carbonate mineralogy and vein-filling materials on Bennu. We therefore searched previously published studies for evidence of carbonate minerals and bright veins in the carbonaceous chondrite meteorites.

The carbonate mineralogy of carbonaceous chondrites is a function of their alteration history. CM chondrites contain on average <4 vol % carbonate (39–41). The least altered CM chondrites contain calcite as the dominant carbonate mineral (39, 42). With increasing degrees of alteration, dolomite appears, followed by breunnerite (42–44). In the most altered samples, dolomite, breunnerite, and calcite are replaced (in that order) by Fe-rich serpentine, Fe–Ni sulfides, and Mg-rich serpentine (39). In CM chondrites of all subtypes, a second generation of calcite forms after the sulfides and phyllosilicates by replacement of remaining anhydrous silicates and dolomite (39).

The carbonate mineralogy of the CI chondrites is similar to that of the most altered CMs. CI meteorites contain four types of carbonates (dolomite, breunnerite, calcite, and siderite) comprising an average of 5 vol % of the samples (24). However, the abundance of carbonates in some individual lithic fragments of CIs exceeds 10 vol %. A lithic fragment in Ivuna (a CI) has calcite as the only carbonate, which accounts for more than 20 vol % (24). Another clast has 21.5 vol % carbonate (45). A dark matrix clast in Orgueuil, another CI, contains ~ 35 vol % of pure calcite (46). CIs have greater diversity in carbonate cations than CMs, including more Mg in breunnerite and magnesite. In CIs, dolomite is the most abundant carbonate phase, with breunnerite and calcite being much less abundant (24, 44, 47). Siderite is extremely rare, with only one grain reported (24).

Similar to the highly altered CMs and the CIs, the ungrouped carbonaceous chondrite Tagish Lake has a carbonate-rich lithology with calcite, dolomite, and Fe- and Mg-rich breunnerite spread diffusely through the matrix (48). The CR chondrites also have carbonates up to 5.7 vol % (49).

Most carbonates in CI and CM chondrites occur as grains 10 to 50 μm in diameter distributed throughout the matrix. However, carbonates in some carbonaceous chondrites occur as veins or vein fragments (39, 50). For

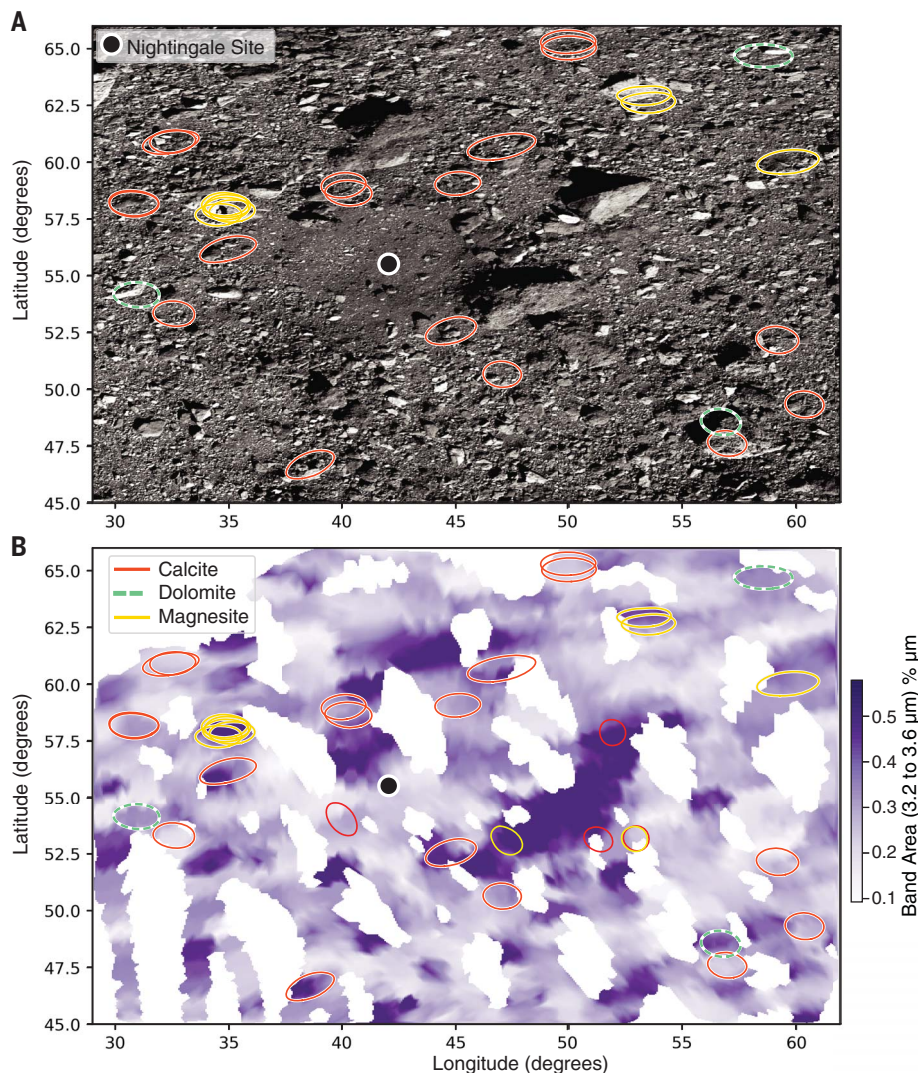


Fig. 4. Maps of carbonate-like spectral signatures at the Nightingale site and surrounding area. OVIRS footprints [colored ellipses, as indicated in the legend in (B)] of differing carbonate compositions are plotted over a panchromatic image (A) and over a map of band area from 3.2 to 3.6 μm (B). The band area does not distinguish between organics and carbonates but demonstrates overall contribution to the spectrum from carbon-bearing species. The Nightingale sample collection site is indicated with a black dot.

instance, the CM chondrite SCO 06043 is crosscut by dolomite veins hundreds of micrometers in length (39). LON 94101, another CM chondrite, contains calcite that occurs as millimeter-sized veins (50). Carbonate aggregates (~100 μm across) resembling fracture fillings have been observed in the Orgueil and Alais CI chondrites (24, 51). These have been interpreted as vein fragments that were mechanically disrupted and dispersed during breakup of the host rock (24). Veins up to ~300 μm in length are observed in Belgica 7904, a chondrite with similarities to both the CI and CM groups (52). Veins of carbonates also occur in some CR chondrites, such as in Renazzo and GRO 95577 (40, 53).

Other vein-filling minerals are found in meteorites. Fractures ≤ 500 μm long are filled with

smectite-chlorite in the ungrouped carbonaceous chondrite Y-82162 (54). Phyllosilicate-magnetite structures and “trains” of magnetite crystals in Orgueil have been interpreted as recrystallized veins (55). Sulfate and phyllosilicate veins occur in some CIs (44), and sulfate and Fe-rich veins are found in CMs (56). At least some of the sulfate veins are thought to be terrestrial weathering products (57).

The presence of bright, carbonate-filled veins in meteorites suggests a connection between our observations of carbonate-matching spectra and bright veins in the Nightingale area. The carbonate compositions that best match the spectral signatures on Bennu are similar to those of the CI chondrites and the most aqueously altered CM chondrites (24, 39, 43, 58), which are sometimes expressed as veins.

Composition of veins on Bennu

In addition to the spectral evidence indicative of carbonates (calcite, magnesite, dolomite, and breunnerite) that we present, known phases on Bennu’s surface include Mg-rich phyllosilicates (3), the Fe-oxide magnetite (3, 4), aliphatic organic molecules (18), and a few pyroxene-rich boulders likely originating from the Vesta asteroid family and delivered by impact (59). We consider whether any of these other phases could be potential vein-filling minerals.

Phyllosilicate veins in meteorites have similar compositions to that of their host rock (54) and therefore are unlikely to have distinct albedos from that of their surroundings, which is inconsistent with the bright veins and irregular features that we observe in host boulders. The exogenous pyroxene-bearing boulders have brighter normal albedos of ~10 to at least 26% (59). However, they have distinctive spectral features in the 1- and 2- μm region that we do not observe in OVIRS spectra covering the bright features. Sulfates are present in Bennu-analog meteorites, but the signatures observed in meteorite spectra (i.e., Ca-sulfate: distinctive features at ~1.4, 1.75, 1.9, and 2.1 to 2.2 μm) (60) are not detected in the OVIRS spectra. Magnetite and organics are optically opaque and unlikely to have the normal albedos of 10 to at least 19% observed for the bright features and veins (23, 61). Thus, although the 3.4- μm feature in some OVIRS spectra could be attributable to organics rather than carbonates (Fig. 2), the low albedo of organic-rich material means that it is not a plausible candidate for the dominant composition of bright veins on Bennu.

We conclude that carbonate is the most likely bright vein-filling material. Magnetite, sulfides, and organic phases may also be present in smaller abundances. These phases are commonly associated with carbonates in carbonaceous chondrites (45). Pure carbonates have albedos approaching 100% (62), but minor abundances of opaque minerals (e.g., magnetite and sulfides) can reduce albedo to the level of the bright features on Bennu (63). Coarser grain size can also reduce albedo (64).

Reciprocally, the bright veins and irregular high-reflectance features in boulders on Bennu are the most likely source of the carbonate-matching spectral signatures that we detected in the higher-spatial resolution OVIRS data. Hereafter, we refer to these 3.4- μm spectral signatures as carbonate signatures.

However, even at the highest spatial resolution available, OVIRS footprints are too large to isolate individual bright features. For instance, the veins in Fig. 6 only account for <0.1% of an OVIRS footprint (~28 m^2), and their host boulders account for 7 to 35% of a

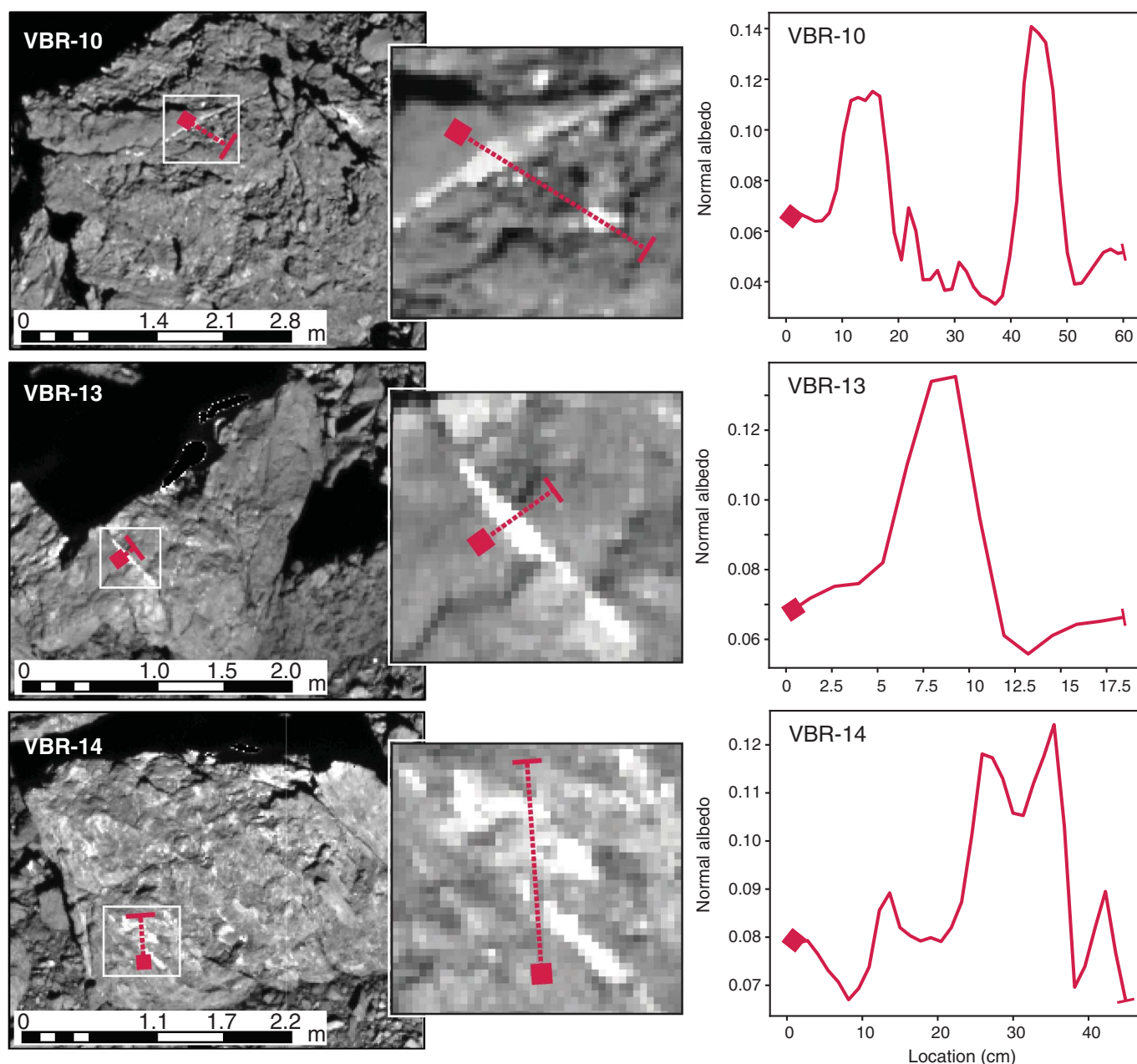


Fig. 5. Images and albedos of bright veins and irregularly shaped features on Bennu. Photometrically corrected PolyCam images (see text) of three boulders in the Nightingale region that host bright veins and irregular features. Transects across the bright features are indicated with red dotted lines, and normal albedos along the transects are plotted in the panels on the right. Albedos exceed 10%. VBR-10 is located at 51.56°N, 58.34°E; VBR-13 at 61.60°N, 50.57°E; and VBR-14 at 62.88°N, 71.55°E. Further images of the VBR-13 are shown in fig. S4.

footprint. Previous work has shown that OVIRS is highly sensitive to bright material, even in cases where it accounts for <1% of the instrument field of view (59). Changes in viewing geometries and outcroppings could block carbonate from the view of the spectrometer, explaining the variable strength of the carbonate feature in the spectral data (fig. S4).

We used a linear mixture analysis to assess whether carbonates, sequestered in veins comprising <0.1% of an OVIRS footprint, could

account for the spectral signatures that we observe (19). We found that very small fractions of carbonate, typically <0.5%, are enough to reproduce the features seen in the OVIRS spectrum (Fig. 6).

Near the Nightingale site, most of the OVIRS spots that cover veins (Fig. 6) have a calcite signature, which is the most common carbonate spectral signature on Bennu (Figs. 2 and 3). Signatures that may indicate more Mg-rich carbonates (dolomite, magnesite, breunnerite) are commonly associated with

larger, bright boulders (up to ~10 m), sometimes without visible evidence for veins or mottling (Fig. 4). Some areas have carbonate signatures with no visible veins or bright irregular features, but where small boulders consistent with the vein-hosting type are present. We attribute these detections not associated with a visible feature to more diffuse carbonate material, as would be expected in boulders or regolith where carbonate deposition may have occurred in micropores or fine veinlets, veins may have been brecciated

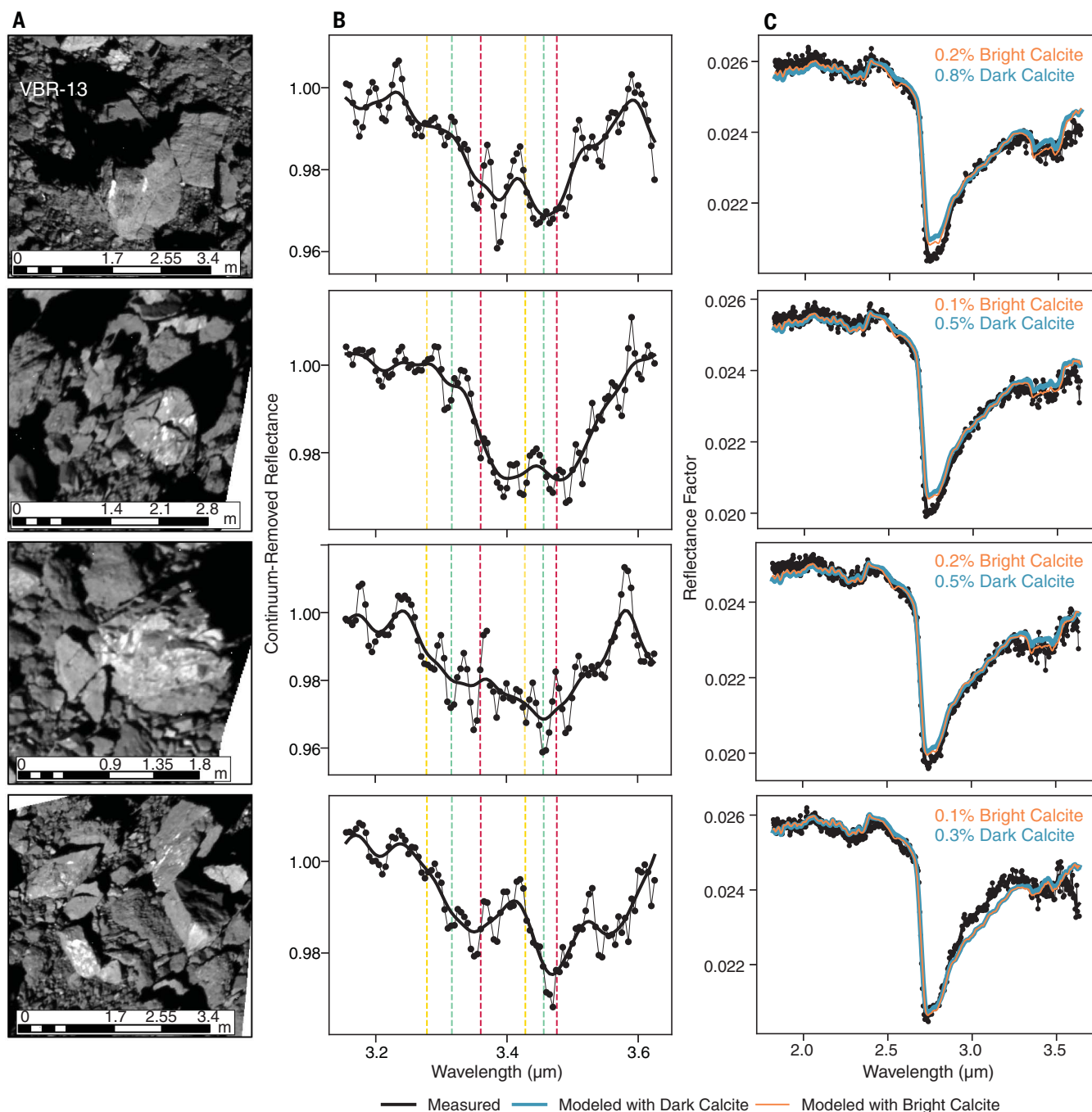


Fig. 6. Boulders with bright features and corresponding spectra. (A) PolyCam images of boulders with bright veins and irregularly shaped features in the Nightingale region. (B) OVIRS spectra of footprints that include each of these boulders have features near $3.4\ \mu\text{m}$ that may be attributable to carbonates. Dashed vertical lines indicate the positions of band minima for calcite (red), dolomite/breunnerite (green), and magnesite (yellow). (C) Linear mixture models (orange and blue lines) reproduce the observations (black data points) using a small fraction ($<1\%$) of carbonate. The extent of the OVIRS footprints is shown in fig. S6, and spectral model endmembers are in fig. S7.

and dispersed through impact gardening, or carbonates may have formed by fine-scale alteration processes that did not cause development of veins. It is also plausible that there is visible-scale carbonate-rich material within a host rock that is simply not exposed on the boulder face (e.g., VBR-14 in Fig. 5).

Conversely, we observe some veins in OVIRS spots without a strong carbonate signature. These spots tend to have spectral signatures of organics or organic-carbonate mixtures, which could indicate that a greater proportion of organics is present that is obscuring the carbonate signature, or that viewing geom-

etry effects are depressing the carbonate band depth (fig. S5) (19).

Carbonate abundances on Bennu

Though carbonaceous chondrites contain carbonate, they rarely show spectral evidence of it in the visible–near-infrared (VIS–NIR)

wavelengths covered by OVIRS (0.4 to 4.3 μm). Carbonaceous chondrites with some of the highest carbonate abundances exhibit a 4- μm carbonate absorption [e.g., Tagish Lake, (23)], but evidence of carbonate in the region from 3.2 to 3.6 μm is rarely found in meteorite spectra. Among published carbonaceous chondrite spectra, <2% have evidence for carbonate in this spectral region, with most either having no absorption or an organic absorption feature (table S4). There are no known organic-carbonate mixtures in carbonaceous chondrite spectra, suggesting that the abundances and/or spectral mixing regime on Bennu are distinct from that in the meteorites. Outcroppings of carbonate in meter-length veins could produce such a distinct spectral mixing regime, which would not be represented in meteorites.

Because carbonate has not been definitively detected in OTES spectra, we can place constraints on its abundance. For particles that are larger than the wavelength of light, thermal infrared spectra are the linear sum of the spectral signatures of all component phases in proportion to their abundance (65). Typically, such data are sensitive to individual phases at volumetric abundances of a few percent or more in laboratory and remote-sensing mixtures (66). The absence of these features suggests that, at global to regional scales, carbonate is unlikely to constitute more than a few percent of Bennu's surface mineralogy. At local scales (single to a few spectra), OTES's SNR is low at wavelengths less than ~7.4 μm , making detection of the ν_3 fundamental band challenging. At longer wavelengths, there are no substantial differences from global and regional averages in the features near 11.3 and 25 to 33 μm . This is consistent with absolute abundances of a few percent or less at the scales of most OTES global observations.

If carbonate is concentrated in the veins and irregular bright features on Bennu, it would account for only a percent or less of the area of each OVIRS spot, or 0.5% of an OTES spot (fig. S6). Such an abundance is consistent with the lack of OTES detection and with our linear mixture analysis, which indicates that only a fraction of a percent of the surface needs to be carbonate to account for the OVIRS signatures (Fig. 6). OVIRS may be particularly sensitive to carbonates if they are segregated from other surface materials and bright compared to their surroundings, as is the case for the pyroxene-bearing boulders (59). Although outcroppings of carbonate at the scales of veins observed on Bennu are not represented in the meteorite collection, the overall abundances of carbonate on Bennu may be similar to those of the meteorites (a few vol % or less). The clasts with high vol % carbonate found in some CI chondrites could result from the breakup of larger features

by impact. We expect high concentrations of carbonate in meteorites if centimeter- and meter-scale carbonate features are present on their parent bodies.

Hydrothermal alteration on Bennu's parent body Distance

Uninterrupted veins up to 150 cm long, such as those that we observe in the OCAMS data, are too large to have formed on Bennu itself and thus must have survived from the parent body. We interpret these veins as most likely originating in the same aqueous system that is responsible for the overall level of hydration observed on Bennu.

Models of carbonaceous asteroid thermal evolution, based on meteoritic evidence, indicate that aqueous alteration was driven by hydrothermal convection [e.g., (67)]. It is unclear whether aqueous alteration occurred in a chemically closed system, in which bulk chemistry did not vary on the length scales of the meteorites, or an open system in which fluid flow promoted chemical fractionation (50). The unfractionated bulk chemical compositions of the carbonaceous chondrites appear incompatible with fluid flow over the large distances expected for open systems (13, 68, 69). Closed-system models are consistent with carbonaceous chondrite unfractionated bulk chemical compositions and their oxygen isotopes (14, 47, 67, 70, 71). The latter models are supported by the low permeability of parent body interiors (72). Open-system hydrothermal alteration scenarios suggest that fluids flowed through fractures from the interior to the exterior of the asteroid, and carbonate minerals precipitated into these fractures as veins (24). Oxygen isotope evidence has been alternatively interpreted as indicative of such open systems with large-scale fluid flow (73, 74). Numerical simulations of parent body evolution also suggest asteroid-scale fluid mobility (75–78).

The putative carbonate veins that we observe on Bennu are orders of magnitude larger than any known in meteorites. We used mass balance constraints to estimate the size of the hydrothermal systems required to produce the observed veins on Bennu. Assuming that the veins are pure calcite, they would contain 56 wt % CaO; this is an upper limit, given the lower albedo of the veins relative to pure carbonate. Previous analysis showed that the bulk abundance of CaO in Orgueil, which has major element abundances similar to the solar photosphere and hence the bulk Solar System, is 1.36 wt %, of which 0.63 wt % is soluble (46). Thus, formation of calcite veins requires a rock volume 41 times that of the vein, if all of the Ca is removed from an Orgueil-like host rock. If only the soluble component is mobilized, then the alteration

volume increases to 89 times that of the vein. The largest observed vein on Bennu is 14 cm thick. This thickness implies that a volume of host rock extending between 288 and 622 cm above and below the vein must have been leached of Ca, if carbonate fills the vein. Thus, hydrothermal alteration on Bennu would have occurred on the scale of at least meters. However, this estimate is a lower limit, as it requires substantial elemental fractionation of nearby host rocks, which is inconsistent with bulk chondritic elemental abundances.

Carbonate formation during aqueous alteration in CM chondrites would have occurred at moderate temperatures (<35°C) and pressures (<10 MPa) in the interior of a parent asteroid (69). These conditions are consistent with experiments of calcite vein formation performed at temperatures of 22° to 23°C and ambient atmospheric pressure (~100 kPa) (79). These experiments indicated that a large volume of fluid is required to produce calcite veins: typically, a fluid/calcite volume ratio of 10^5 to 10^6 for formation of thin short veins. Applying these values to the thickest veins on Bennu suggests fluid flow at the scale of 20 to 45 km on either side of the vein, implying a nearly global, open hydrologic system on Bennu's ~100-km-diameter parent body (11, 19).

Our calculations support the open-system model of asteroidal aqueous alteration. In this scenario, internal pressurization drove fluid flow from the interior of Bennu's parent asteroid through pore spaces to the asteroid surface, where water could be lost into space. As water ice melted, liquid water migrated through high-permeability fractures created by impact-induced brecciation (24, 39, 50, 58, 80). Alternatively, crack propagation could have been driven by overpressurization of gas formed during alteration (81). The reactants Ca and CO₂ were leached from host rock and transported over distances of meters to kilometers. As the hydrothermal solution cooled, calcite crystals grew into open pore spaces along the fractures, forming the veins that we observe on Bennu. This mechanism is consistent with calcite in CM chondrites having formed by cementation of fluid-filled fractures (50, 58).

Impact processing of the parent asteroid regolith could have continued after vein formation. The catastrophic disruption of the parent body and reaccumulation of rubble to form Bennu likely fractured rocks along the weak carbonate veins, potentially explaining why they are sometimes exposed on apparently broken faces. Thus, the geologic expression of the veins on Bennu has been modified by its violent history.

Time

Aqueous alteration likely proceeded progressively in multiple phases or events. Several

episodes of alteration occurred on the CI parent body (24). Coexisting calcites and dolomites in CM chondrites are not in equilibrium so also formed in distinct events (44, 80). Dolomites typically form slightly earlier than breunnerites. Extensive aqueous activity on the Orgueil parent body occurred over a period of several million years, starting ~3 to 4 million years (Myr) after formation of the Solar System (82). The probable difference in age between dolomite and breunnerite indicates temporal variations in fluid composition (82). Different carbonates (e.g., calcite, dolomite, and breunnerite) are characterized by distinct oxygen and carbon isotopic compositions [e.g., (83)]. Much of the calcite in CM chondrites represents a second generation that formed after the first-stage alteration that established the dominant phyllosilicate mineralogy [e.g., (39)]. This scenario is consistent with our observations of bright, ostensibly calcite veins hosted in fractures in phyllosilicate-rich boulders.

Carbonate precipitation time scales depend on factors such as fluid saturation state, flow velocity, vein dimension, and kinetic rate constants and reaction order. We apply a kinetic model (79) to constrain the time necessary for vein filling, finding that the veins on Benu developed on the parent body over thousands to millions of years (19).

These time scales are consistent with the timing of crystallization indicated by Mn-Cr dating of carbonates in CI and CM chondrites. Calcites and dolomites in four CM chondrites all have formation ages of $3.8^{+0.4}_{-0.5}$ Myr after the formation of the oldest Solar System solids (84). The crystallization ages of dolomite in the QUE 93005 and Sutter's Mill carbonaceous chondrites indicate that aqueous alteration occurred at a similar timing of 3.93 ± 0.23 Myr (42) and 2.4 to 5.0 Myr after the birth of the Solar System (85), respectively.

Does Benu have a meteorite analog?

Spectral data from OVIRS and OTES collected while approaching the asteroid indicated that Benu's mineralogy and chemistry were consistent with the most aqueously altered CM chondrites but did not rule out the presence of some CI-like material (3). We build on that analysis with the spectral and geologic features that we observe at Nightingale to further constrain Benu's relationship to the carbonaceous chondrite meteorites.

Benu's observed global spectral properties are broadly consistent with a highly altered CM chondrite mineralogy (e.g., CM1): Benu has a hydration band minimum at $2.74 \mu\text{m}$, a position that is associated primarily with Mg-serpentine (e.g., antigorite) but could incorporate some Fe-serpentine (cronstedtite) (3). The CIs, however, have band minima at shorter wavelengths ($2.71 \mu\text{m}$) associated with the phyllosilicates saponite and Mg-serpentine

(lizardite and chrysotile) (6). The OTES data suggest that Benu's surface is dominated by phyllosilicates, consistent with the most aqueously altered CMs (3).

Though calcite is the most common carbonate mineral that we identify on Benu, at least one-quarter of the carbonate detections are more Mg-rich compositions based on the wavelength of their $3.4\text{-}\mu\text{m}$ features (Fig. 3), including likely dolomite, breunnerite, and magnesite. Only calcite, breunnerite, and dolomite have been found in CM chondrites, possibly because Mg is limited to dissolution from olivine, which occurs in only the most altered CMs (39, 43). Magnesite has only been reported in the CIs (47, 86) and in some ungrouped carbonaceous chondrites (87), so the presence of magnesite on Benu could imply a closer relationship to these meteorites. However, calcites still account for the largest individual fragments in the CIs (46), and vein-filling calcites are found in both CIs and CMs (24, 39, 51).

CM and CI meteorites are usually assumed to originate from different asteroid parent bodies. The bright veins and irregular features on Benu are associated with a specific boulder morphology and reflectance range, so they could have a different composition, and thus a distinct source, from that of the other boulders on Benu. However, though the vein-bearing boulders belong to a population with distinctive albedo and color (37), there is no indication that the hydration band minimum position or depth for these boulders is different from that of the rest of Benu (18), though much of the spectral data have a spatial resolution too low to isolate individual boulders of this type. Benu's rubble-pile structure and diverse boulder morphologies are consistent with a mixture of compositions from multiple locations within its parent asteroid.

Benu may bear the closest resemblance to the ungrouped carbonaceous chondrite meteorites with similarities to both CMs and CIs. For instance, the ungrouped carbonaceous chondrite Tagish Lake has mineralogy, oxygen isotope, and bulk chemical composition that fall between those of the CMs and CIs (88). It has a carbonate-rich lithology that contains both calcite and Fe,Mg-carbonates intermixed with saponite and magnetite clasts (88). Tagish Lake is dynamically and spectrally linked to outer main belt C-complex asteroids with red spectral slopes (89), so we do not suggest that it is directly related to the blue-sloped Benu. However, Benu may similarly represent an ungrouped, intermediate composition of chondritic material. The bright centimeter-scale irregular features and meter-scale veins on Benu, which are also not represented in meteorites, may be characteristic of this particular material; alternatively, it may be common on carbonaceous

chondrite parent bodies but only observed on Benu, where we are able to make observations at the outcrop scale. Mottled boulders on Ryugu could be a similar expression of bright, carbonate-bearing material; however, this cannot be confirmed with the spectrometer on Hayabusa2, which does not cover the $3.4\text{-}\mu\text{m}$ region.

Vein fragments in carbonaceous chondrites could be rare because most carbonate veins have been broken up and scattered by impacts (44, 51). Carbonate veins may be preferentially destroyed by the ejection process and/or passage through Earth's atmosphere, reducing the likelihood of being represented in the meteorite record. Meteorite collection parties might not recognize carbonate-rich fragments as possible meteorites.

Role of space weathering

Space weathering can modify the spectrum of an exposed surface, complicating comparison between in situ spectra of Benu and the interiors of meteorites. The influence of space weathering on Benu's spectral properties is only partly understood. The bright boulder population that hosts the veins and irregular features appears more resistant to space weathering than dark boulders on Benu (37). Irradiation experiments performed on meteorites indicate that a wide range of changes in VIS-NIR spectral slope are possible (90). Space weathering can destroy aliphatic organics (91), which could explain the variable organic and carbonate mixture signatures that we observe with OVIRS. In this scenario, more recently exposed surfaces would have stronger organic absorptions that weaken with time. Benu is known to eject particles from its surface; the mechanism is unknown but could expose material from the interior (92). The spectral signature of carbonate and mixtures, where present, still implies either an unexpectedly high abundance or, more likely, an unusual outcropping of carbonate, compared to the carbonaceous chondrite spectra.

OSIRIS-REx's sampling mechanism is designed to collect the top several centimeters of Benu's regolith (93), whereas the spectrometers only measure the top tens to hundreds of micrometers. Thus, space-weathered material may be in the minority of the collected sample, unless regolith mixes on the centimeter scale on time scales longer than the rate of space weathering. Material attached to the spacecraft's surface contact pads could distinguish surface-exposed material from un-space-weathered material in the near subsurface.

Thermal evolution and organics

Most aqueous alteration in the CIs and CMs occurred at temperatures $<150^\circ\text{C}$, but some meteorites have experienced additional heating to higher temperatures (86). Carbonates

decompose at ~400° to 800°C depending on cation (94), suggesting that the carbonates on Bennu were not exposed to these higher temperatures, at least not for a sufficiently long period to decompose them all. The pervasive 2.7-μm hydration feature, with no spectral evidence for volumetrically abundant anhydrous silicates such as olivine and pyroxene (3), indicates that the phyllosilicate that is widespread on Bennu has not been dehydrated or decomposed (which is nearly complete by 700°C), nor has it formed detectable secondary olivine (which occurs by 800°C) (95).

The temperature and duration of aqueous alteration of the materials on Bennu, and their thermal metamorphism, will affect the abundance and distribution of both soluble and insoluble organic matter in the returned samples. The presence of aliphatic organics across Bennu's surface places a stricter limit on peak temperatures than the carbonates, because thermal alteration quickly destroys aliphatic bonds (96). The 3.4-μm spectral feature from aliphatic organics in insoluble organic matter (IOM) is lost within ~200 years at 100°C (97). CMs with organic spectral features therefore experienced <15°C heating over the duration of aqueous alteration, potentially further limiting the range of temperatures experienced by organic-rich regions on Bennu. It is possible that organic-rich boulders and carbonate-rich boulders are distinct populations (40, 41), with different thermal histories. The thermal gradient within an asteroid parent body similar to Bennu's spans a range of hundreds of degrees Celsius (73), and this gradient, along with spatial heterogeneity in pore water flow, could account for the range of compositions inherited by Bennu.

Conclusions

Remote-sensing observations provide outcrop-scale information, on scales larger than meteorite samples. Bennu shares compositional traits with aqueously altered CM and CI meteorites, including the presence of Mg-phyllosilicates, magnetite, and, we posit, carbonate mineralogy dominated by calcite, with smaller amounts of magnesite, dolomite, and breunnerite. However, the carbonate spectral dominance in OVIRS data at 3.4 μm is unusual compared to meteorite spectra. We observe bright veins with thicknesses of 3 to 15 cm and lengths that can exceed 1 m. We propose that they are composed of the spectrally detected carbonates. On the basis of the vein dimensions, which are much larger than any veins found in meteorites, a hydrothermal system of at least 89 times the volume of the veins, or an amount of fluid at least 10⁵ times the volume of the veins, would have been required to leach the elements from the surrounding host rock. Fluid flow on Bennu's parent body would have taken place over distances of kilometers

for thousands to millions of years. We predict that the returned sample could contain carbonates with structure and scale distinct from those in the meteorites.

It remains unclear whether these differences arise from material that would not survive delivery to Earth as a meteorite, or because Bennu's parent body is distinct from the CM or CI parent body. Bennu is unlikely to be the only object of its kind in the asteroid population, as most asteroids have not been visited by spacecraft. Large-scale open-system hydrothermal alteration may therefore have occurred on carbonaceous parent bodies more generally. Ceres, the only other known carbonate-bearing asteroid that has been visited by a spacecraft, experienced extensive hydrothermal alteration, beyond the degree of the most hydrated (CI group) meteorites (98). This extensive alteration along with more recent geologic activity (99) may have obscured evidence of the earliest stages of alteration, including vein formation. Whether distinct or representative, the returned sample from Bennu is likely to contain carbon-bearing species in the form of organics and/or carbonates.

REFERENCES AND NOTES

1. D. S. Lauretta *et al.*, OSIRIS-REx: Sample Return from Asteroid (101955) Bennu. *Space Sci. Rev.* **212**, 925–984 (2017). doi: [10.1007/s11214-017-0405-1](https://doi.org/10.1007/s11214-017-0405-1)
2. B. E. Clark *et al.*, Asteroid (101955) 1999 RQ36: Spectroscopy from 0.4 to 2.4 μm and meteorite analogs. *Icarus* **216**, 462–475 (2011). doi: [10.1016/j.icarus.2011.08.021](https://doi.org/10.1016/j.icarus.2011.08.021)
3. V. E. Hamilton *et al.*, Evidence for widespread hydrated minerals on asteroid (101955) Bennu. *Nat. Astron.* **3**, 332–340 (2019). doi: [10.1038/s41550-019-0722-2](https://doi.org/10.1038/s41550-019-0722-2); pmid: [31360777](https://pubmed.ncbi.nlm.nih.gov/31360777/)
4. D. S. Lauretta *et al.*, The unexpected surface of asteroid (101955) Bennu. *Nature* **568**, 55–60 (2019). doi: [10.1038/s41586-019-1033-6](https://doi.org/10.1038/s41586-019-1033-6); pmid: [30890786](https://pubmed.ncbi.nlm.nih.gov/30890786/)
5. D. DellaGiustina *et al.*, Properties of rubble-pile asteroid (101955) Bennu from OSIRIS-REx imaging and thermal analysis. *New Astron.* **3**, 341–351 (2019).
6. D. Takir *et al.*, Nature and degree of aqueous alteration in CM and CI carbonaceous chondrites. *Meteorit. Planet. Sci.* **48**, 1618–1637 (2013). doi: [10.1111/maps.12171](https://doi.org/10.1111/maps.12171)
7. A. Ghosh, S. J. Weidenschilling, H. Y. McSweeney Jr., A. E. Rubin, in *Meteorites and the Early Solar System II* (Univ. of Arizona Press, 2006), pp. 555–566.
8. M. Endress, E. Zinner, A. Bischoff, Early aqueous activity on primitive meteorite parent bodies. *Nature* **379**, 701–703 (1996). doi: [10.1038/379701a0](https://doi.org/10.1038/379701a0); pmid: [8602215](https://pubmed.ncbi.nlm.nih.gov/8602215/)
9. O. S. Barnouin *et al.*, Shape of (101955) Bennu indicative of a rubble pile with internal stiffness. *Nat. Geosci.* **12**, 247–252 (2019). doi: [10.1038/s41561-019-0330-x](https://doi.org/10.1038/s41561-019-0330-x); pmid: [31080497](https://pubmed.ncbi.nlm.nih.gov/31080497/)
10. P. Michel *et al.*, Collisional formation of top-shaped asteroids and implications for the origins of Ryugu and Bennu. *Nat. Commun.* **11**, 2655 (2020). doi: [10.1038/s41467-020-16433-z](https://doi.org/10.1038/s41467-020-16433-z); pmid: [32461569](https://pubmed.ncbi.nlm.nih.gov/32461569/)
11. K. J. Walsh, M. Delbó, W. F. Bottke, D. Vokrouhlický, D. S. Lauretta, Introducing the Eulalia and new Polana asteroid families: Re-assessing primitive asteroid families in the inner Main Belt. *Icarus* **225**, 283–297 (2013). doi: [10.1016/j.icarus.2013.03.005](https://doi.org/10.1016/j.icarus.2013.03.005)
12. W. R. Van Schmus, J. A. Wood, A chemical-petrologic classification for the chondritic meteorites. *Geochim. Cosmochim. Acta* **31**, 747–765 (1967). doi: [10.1016/S0016-7037\(67\)80030-9](https://doi.org/10.1016/S0016-7037(67)80030-9)
13. H. Y. McSweeney Jr., S. M. Richardson, The composition of carbonaceous chondrite matrix. *Geochim. Cosmochim. Acta* **41**, 1145–1161 (1977). doi: [10.1016/0016-7037\(77\)90110-7](https://doi.org/10.1016/0016-7037(77)90110-7)
14. A. E. Rubin, J. M. Trigo-Rodríguez, H. Huber, J. T. Wasson, Progressive aqueous alteration of CM carbonaceous chondrites. *Geochim. Cosmochim. Acta* **71**, 2361–2382 (2007). doi: [10.1016/j.gca.2007.02.008](https://doi.org/10.1016/j.gca.2007.02.008)
15. M. E. Zolensky *et al.*, D. B. CM chondrites exhibit the complete petrologic range from type 2 to 1. *Geochim. Cosmochim. Acta* **61**, 5099–5115 (1997). doi: [10.1016/S0016-7037\(97\)00357-8](https://doi.org/10.1016/S0016-7037(97)00357-8)
16. G. W. Kallemeyn, J. T. Wasson, The compositional classification of chondrites: IV. Ungrouped chondritic meteorites and clasts. *Geochim. Cosmochim. Acta* **49**, 261–270 (1985). doi: [10.1016/0016-7037\(85\)90209-1](https://doi.org/10.1016/0016-7037(85)90209-1)
17. D. C. Reuter *et al.*, The OSIRIS-REx Visible and InfraRed Spectrometer (OVIRS): Spectral Maps of the Asteroid Bennu. *Space Sci. Rev.* **214**, 54 (2018). doi: [10.1007/s11214-018-0482-9](https://doi.org/10.1007/s11214-018-0482-9)
18. A. A. Simon *et al.*, Widespread carbon-bearing materials on near-Earth asteroid (101955) Bennu. *Science* **370**, eabc3522 (2020). doi: [10.1126/science.abc3522](https://doi.org/10.1126/science.abc3522)
19. Materials and methods are available as supplementary materials.
20. M. C. De Sanctis *et al.*, Localized aliphatic organic material on the surface of Ceres. *Science* **355**, 719–722 (2017). doi: [10.1126/science.aaj2305](https://doi.org/10.1126/science.aaj2305); pmid: [28209893](https://pubmed.ncbi.nlm.nih.gov/28209893/)
21. A. Raponi *et al.*, Infrared detection of aliphatic organics on a cometary nucleus. *New Astron.* **4**, 500–505 (2020).
22. A. S. Rivkin, J. P. Emery, Detection of ice and organics on an asteroidal surface. *Nature* **464**, 1322–1323 (2010). doi: [10.1038/nature09028](https://doi.org/10.1038/nature09028); pmid: [20428165](https://pubmed.ncbi.nlm.nih.gov/20428165/)
23. H. H. Kaplan, R. E. Milliken, C. M. O. Alexander, C. D. K. Herd, Reflectance spectroscopy of insoluble organic matter (IOM) and carbonaceous meteorites. *Meteorit. Planet. Sci.* **54**, 1051–1068 (2019). doi: [10.1111/maps.13264](https://doi.org/10.1111/maps.13264)
24. M. Endress, A. Bischoff, Carbonates in CI chondrites: Clues to parent body evolution. *Geochim. Cosmochim. Acta* **60**, 489–507 (1996). doi: [10.1016/0016-7037\(95\)00399-1](https://doi.org/10.1016/0016-7037(95)00399-1); pmid: [11539921](https://pubmed.ncbi.nlm.nih.gov/11539921/)
25. S. J. Gaffey, Spectral reflectance of carbonate minerals in the visible and near infrared (0.35–2.55 μm): Anhydrous carbonate minerals. *J. Geophys. Res.* **92** (B2), 1429 (1987). doi: [10.1029/JB092B02p01429](https://doi.org/10.1029/JB092B02p01429)
26. M. C. De Sanctis *et al.*, Bright carbonate deposits as evidence of aqueous alteration on (1) Ceres. *Nature* **536**, 54–57 (2016). doi: [10.1038/nature18290](https://doi.org/10.1038/nature18290); pmid: [27362221](https://pubmed.ncbi.nlm.nih.gov/27362221/)
27. P. R. Christensen *et al.*, The OSIRIS-REx Thermal Emission Spectrometer (OTES) Instrument. *Space Sci. Rev.* **214**, 87 (2018). doi: [10.1007/s11214-018-0513-6](https://doi.org/10.1007/s11214-018-0513-6)
28. J. L. Bandfield, T. D. Glotch, P. R. Christensen, Spectroscopic identification of carbonate minerals in the martian dust. *Science* **301**, 1084–1087 (2003). doi: [10.1126/science.1088054](https://doi.org/10.1126/science.1088054); pmid: [12934004](https://pubmed.ncbi.nlm.nih.gov/12934004/)
29. K. C. Feely, P. R. Christensen, Quantitative compositional analysis using thermal emission spectroscopy: Application to igneous and metamorphic rocks. *J. Geophys. Res. Planets* **104** (E10), 24195–24210 (1999). doi: [10.1029/1999JE001034](https://doi.org/10.1029/1999JE001034)
30. B. Rizk *et al.*, OCAMS: The OSIRIS-REx Camera Suite. *Space Sci. Rev.* **214**, 26 (2018). doi: [10.1007/s11214-017-0460-7](https://doi.org/10.1007/s11214-017-0460-7)
31. D. R. Golish *et al.*, Ground and In-Flight Calibration of the OSIRIS-REx Camera Suite. *Space Sci. Rev.* **216**, 12 (2020). doi: [10.1007/s11214-019-0626-6](https://doi.org/10.1007/s11214-019-0626-6); pmid: [32025061](https://pubmed.ncbi.nlm.nih.gov/32025061/)
32. O. S. Barnouin *et al.*, Digital terrain mapping by the OSIRIS-REx mission. *Planet. Space Sci.* **180**, 104764 (2020). doi: [10.1016/j.pss.2019.104764](https://doi.org/10.1016/j.pss.2019.104764)
33. D. R. Golish *et al.*, Disk-resolved photometric modeling and properties of asteroid (101955) Bennu. *Icarus* **10.1016/j.icarus.2020.113724** (2020). doi: [10.1016/j.icarus.2020.113724](https://doi.org/10.1016/j.icarus.2020.113724)
34. C. W. Hergenrother *et al.*, The operational environment and rotational acceleration of asteroid (101955) Bennu from OSIRIS-REx observations. *Nat. Commun.* **10**, 1291 (2019). doi: [10.1038/s41467-019-09213-x](https://doi.org/10.1038/s41467-019-09213-x); pmid: [30890725](https://pubmed.ncbi.nlm.nih.gov/30890725/)
35. T. G. Sharp, P. S. DeCarli, in *Meteorites and the Early Solar System II* (Univ. of Arizona Press, 2006), pp. 653–677.
36. S. Sugita *et al.*, The geomorphology, color, and thermal properties of Ryugu: Implications for parent-body processes. *Science* **364**, 252 (2019). doi: [10.1126/science.aaw0422](https://doi.org/10.1126/science.aaw0422); pmid: [30890587](https://pubmed.ncbi.nlm.nih.gov/30890587/)
37. D. N. DellaGiustina *et al.*, Variation in color and reflectance of asteroid (101955) Bennu. *Science* (2020). doi: [10.1126/science.abc3660](https://doi.org/10.1126/science.abc3660)
38. B. Rozitis *et al.*, Asteroid (101955) Bennu's Weak Boulders and Thermally Anomalous Equator. *Sci. Adv.* (2020). doi: [10.1126/sciadv.abc3699](https://doi.org/10.1126/sciadv.abc3699)
39. M. R. Lee, P. Lindgren, M. R. Sofer, Aragonite, breunnerite, calcite and dolomite in the CM carbonaceous chondrites: High fidelity recorders of progressive parent body aqueous alteration. *Geochim. Cosmochim. Acta* **144**, 126–156 (2014). doi: [10.1016/j.gca.2014.08.019](https://doi.org/10.1016/j.gca.2014.08.019)

40. K. T. Howard, C. M. O. Alexander, D. L. Schrader, K. A. Dyl, Classification of hydrous meteorites (CR, CM and C2 ungrouped) by phyllosilicate fraction: PSD-XRD modal mineralogy and planetesimal environments. *Geochim. Cosmochim. Acta* **149**, 206–222 (2015). doi: [10.1016/j.gca.2014.10.025](https://doi.org/10.1016/j.gca.2014.10.025)
41. C. M. O. Alexander, R. Bowden, M. L. Fogel, K. T. Howard, Carbonate abundances and isotopic compositions in chondrites. *Meteorit. Planet. Sci.* **50**, 810–833 (2015). doi: [10.1111/maps.12410](https://doi.org/10.1111/maps.12410)
42. M. R. Lee, P. Lindgren, M. R. Sofo, C. M. O'D Alexander, J. Wang, C. M. O'D Alexander, J. Wang, Extended chronologies of aqueous alteration in the CM2 carbonaceous chondrites: Evidence from carbonates in Queen Alexandra Range 93005. *Geochim. Cosmochim. Acta* **92**, 148–169 (2012). doi: [10.1016/j.gca.2012.06.005](https://doi.org/10.1016/j.gca.2012.06.005)
43. S. De Leuw, A. E. Rubin, J. T. Wasson, Carbonates in CM chondrites: Complex formation histories and comparison to carbonates in CI chondrites. *Meteorit. Planet. Sci.* **45**, 513–530 (2010). doi: [10.1111/j.1945-5100.2010.01037.x](https://doi.org/10.1111/j.1945-5100.2010.01037.x)
44. L. R. Ricuputi, H. Y. McSween Jr., C. A. Johnson, M. Prinz, Minor and trace element concentrations in carbonates of carbonaceous chondrites, and implications for the compositions of coexisting fluids. *Geochim. Cosmochim. Acta* **58**, 1343–1351 (1994). doi: [10.1016/0016-7037\(94\)90386-7](https://doi.org/10.1016/0016-7037(94)90386-7)
45. J. Alfing, M. Patzek, A. Bischoff, Modal abundances of coarse-grained (>5 µm) components within CI-chondrites and their individual clasts – Mixing of various lithologies on the CI parent body(ies). *Geochemistry* **79**, 125532 (2019). doi: [10.1016/j.chemer.2019.08.004](https://doi.org/10.1016/j.chemer.2019.08.004)
46. K. Fredriksson, J. F. Kerridge, Carbonates and sulfates in CI chondrites: Formation by aqueous activity on the parent body. *Meteoritics* **23**, 35–44 (1988). doi: [10.1111/j.1945-5100.1988.tb00894.x](https://doi.org/10.1111/j.1945-5100.1988.tb00894.x); pmid: [11538410](https://pubmed.ncbi.nlm.nih.gov/11538410/)
47. C. A. Johnson, M. Prinz, Carbonate compositions in CM and CI chondrites and implications for aqueous alteration. *Geochim. Cosmochim. Acta* **57**, 2843–2852 (1993). doi: [10.1016/0016-7037\(93\)90393-B](https://doi.org/10.1016/0016-7037(93)90393-B)
48. T. Nakamura, Yamato 793321 CM chondrite: Dehydrated regolith material of a hydrous asteroid. *Earth Planet. Sci. Lett.* **242**, 26–38 (2006). doi: [10.1016/j.epsl.2005.11.040](https://doi.org/10.1016/j.epsl.2005.11.040)
49. M. K. Weisberg, M. Prinz, R. N. Clayton, T. K. Mayeda, The CR (Renazzo-type) carbonaceous chondrite group and its implications. *Geochim. Cosmochim. Acta* **57**, 1567–1586 (1993). doi: [10.1016/0016-7037\(93\)90013-M](https://doi.org/10.1016/0016-7037(93)90013-M)
50. M. R. Lee, M. R. Sofo, P. Lindgren, N. A. Starkey, I. A. Franchi, The oxygen isotope evolution of parent body aqueous solutions as recorded by multiple carbonate generations in the Lonewolf Nunataks 94101 CM2 carbonaceous chondrite. *Geochim. Cosmochim. Acta* **121**, 452–466 (2013). doi: [10.1016/j.gca.2013.07.010](https://doi.org/10.1016/j.gca.2013.07.010)
51. S. M. Richardson, Vein formation in the CI carbonaceous chondrites. *Meteoritics* **13**, 141–159 (1978). doi: [10.1111/j.1945-5100.1978.tb00803.x](https://doi.org/10.1111/j.1945-5100.1978.tb00803.x)
52. K. Tomeoka, in *Proceedings of the NIPR Symposium*, K. Yanai, Ed. (National Institute of Polar Research, 1990), vol. 3.
53. C. E. Jilly-Rehak, G. R. Huss, K. Nagashima, 53 Mn– 53 Cr radiometric dating of secondary carbonates in CR chondrites: Timescales for parent body aqueous alteration. *Geochim. Cosmochim. Acta* **201**, 224–244 (2017). doi: [10.1016/j.gca.2016.08.033](https://doi.org/10.1016/j.gca.2016.08.033)
54. K. Tomeoka, Phyllosilicate veins in a CI meteorite: Evidence for aqueous alteration on the parent body. *Nature* **345**, 138–140 (1990). doi: [10.1038/345138a0](https://doi.org/10.1038/345138a0)
55. M. R. Lee, K. Nicholson, Ca-carbonate in the Orgueil (CI) carbonaceous chondrite: Mineralogy, microstructure and implications for parent body history. *Earth Planet. Sci. Lett.* **280**, 268–275 (2009). doi: [10.1016/j.epsl.2009.01.038](https://doi.org/10.1016/j.epsl.2009.01.038)
56. N. P. Hanowski, A. J. Brearley, Iron-rich aureoles in the CM carbonaceous chondrites Murray, Murchison, and Allan Hills 81002: Evidence for *in situ* aqueous alteration. *Meteorit. Planet. Sci.* **35**, 1291–1308 (2000). doi: [10.1111/j.1945-5100.2000.tb01517.x](https://doi.org/10.1111/j.1945-5100.2000.tb01517.x)
57. M. Gounelle, M. E. Zolensky, A terrestrial origin for sulfate veins in CI chondrites. *Meteorit. Planet. Sci.* **36**, 1321–1329 (2001). doi: [10.1111/j.1945-5100.2001.tb01827.x](https://doi.org/10.1111/j.1945-5100.2001.tb01827.x)
58. P. Lindgren, M. R. Lee, M. Sofo, M. J. Burchell, Microstructure of calcite in the CM2 carbonaceous chondrite LON 94101: Implications for deformation history during and/or after aqueous alteration. *Earth Planet. Sci. Lett.* **306**, 289–298 (2011). doi: [10.1016/j.epsl.2011.04.022](https://doi.org/10.1016/j.epsl.2011.04.022)
59. D. N. DellaGiustina *et al.*, Exogenic Basalt on Asteroid (101955) Bennu. *New Astron.* **10.1038/s41550-020-1195-z** (2020). doi: [10.1038/s41550-020-1195-z](https://doi.org/10.1038/s41550-020-1195-z)
60. E. A. Cloutis, T. Hiroi, M. J. Gaffey, C. M. O. Alexander, P. Mann, Spectral reflectance properties of carbonaceous chondrites: I. CI chondrites. *Icarus* **212**, 180–209 (2011). doi: [10.1016/j.icarus.2010.12.009](https://doi.org/10.1016/j.icarus.2010.12.009)
61. M. R. M. Izawa *et al.*, Spectral reflectance properties of magnetites: Implications for remote sensing. *Icarus* **319**, 525–539 (2019). doi: [10.1016/j.icarus.2018.10.002](https://doi.org/10.1016/j.icarus.2018.10.002)
62. E. A. Cloutis *et al.*, Spectral reflectance properties of carbonates from terrestrial analogue environments: Implications for Mars. *Planet. Space Sci.* **58**, 522–537 (2010). doi: [10.1016/j.pss.2009.09.002](https://doi.org/10.1016/j.pss.2009.09.002)
63. R. N. Clark, Spectral properties of mixtures of montmorillonite and dark carbon grains: Implications for remote sensing minerals containing chemically and physically adsorbed water. *J. Geophys. Res.* **88**, 10635–10644 (1983). doi: [10.1029/JB088iB12p10635](https://doi.org/10.1029/JB088iB12p10635)
64. A. Le Bras, S. Erard, Reflectance spectra of regolith analogs in the mid-infrared: Effects of grain size. *Planet. Space Sci.* **51**, 281–294 (2003). doi: [10.1016/S0032-0633\(03\)00017-5](https://doi.org/10.1016/S0032-0633(03)00017-5)
65. R. J. P. Lyon, W. M. Tuddenham, C. S. Thompson, Quantitative mineralogy in 30 minutes. *Econ. Geol.* **54**, 1047–1055 (1959). doi: [10.2113/gsecongeo.54.6.1047](https://doi.org/10.2113/gsecongeo.54.6.1047)
66. A. D. Rogers, O. Aharonson, Mineralogical composition of sands in Meridiani Planum determined from Mars Exploration Rover data and comparison to orbital measurements. *J. Geophys. Res.* **113**, E06S14 (2008). doi: [10.1029/2007JE002995](https://doi.org/10.1029/2007JE002995)
67. R. E. Grimm, H. Y. McSween Jr., Water and the thermal evolution of carbonaceous chondrite parent bodies. *Icarus* **82**, 244–280 (1989). doi: [10.1016/0019-1035\(89\)90038-9](https://doi.org/10.1016/0019-1035(89)90038-9)
68. A. J. Brearley, M. Prinz, CI chondrite-like clasts in the Nilpena polymict ureilite: Implications for aqueous alteration processes in CI chondrites. *Geochim. Cosmochim. Acta* **56**, 1373–1386 (1992). doi: [10.1016/0016-7037\(92\)90068-T](https://doi.org/10.1016/0016-7037(92)90068-T)
69. M. Zolensky, R. Barrett, L. Browning, Mineralogy and composition of matrix and chondrule rims in carbonaceous chondrites. *Geochim. Cosmochim. Acta* **57**, 3123–3148 (1993). doi: [10.1016/0016-7037\(93\)90298-B](https://doi.org/10.1016/0016-7037(93)90298-B)
70. E. R. Dufresne, E. Anders, On the chemical evolution of the carbonaceous chondrites. *Geochim. Cosmochim. Acta* **26**, 1085–1114 (1962). doi: [10.1016/0016-7037\(62\)90047-9](https://doi.org/10.1016/0016-7037(62)90047-9)
71. R. N. Clayton, T. K. Mayeda, Oxygen isotope studies of carbonaceous chondrites. *Geochim. Cosmochim. Acta* **63**, 2089–2104 (1999). doi: [10.1016/S0016-7037\(99\)00090-3](https://doi.org/10.1016/S0016-7037(99)00090-3)
72. P. A. Bland *et al.*, Why aqueous alteration in asteroids was isochemical: High porosity/high permeability. *Earth Planet. Sci. Lett.* **287**, 559–568 (2009). doi: [10.1016/j.epsl.2009.09.004](https://doi.org/10.1016/j.epsl.2009.09.004)
73. E. D. Young, R. D. Ash, P. England, D. Rumble 3rd, Fluid flow in chondritic parent bodies: Deciphering the compositions of planetesimals. *Science* **286**, 1331–1335 (1999). doi: [10.1126/science.286.5443.1331](https://doi.org/10.1126/science.286.5443.1331); pmid: [10558982](https://pubmed.ncbi.nlm.nih.gov/10558982/)
74. E. D. Young, The hydrology of carbonaceous chondrite parent bodies and the evolution of planet progenitors. *Philos. Trans. - Royal Soc., Math. Phys. Eng. Sci.* **359**, 2095–2110 (2001). doi: [10.1098/rsta.2001.0900](https://doi.org/10.1098/rsta.2001.0900)
75. B. Travis, G. Schubert, Hydrothermal convection in carbonaceous chondrite parent bodies. *Earth Planet. Sci. Lett.* **240**, 234–250 (2005). doi: [10.1016/j.epsl.2005.09.008](https://doi.org/10.1016/j.epsl.2005.09.008)
76. E. D. Young, K. K. Zhang, G. Schubert, Conditions for pore water convection within carbonaceous chondrite parent bodies – implications for planetesimal size and heat production. *Earth Planet. Sci. Lett.* **213**, 249–259 (2003). doi: [10.1016/S0012-821X\(03\)00345-5](https://doi.org/10.1016/S0012-821X(03)00345-5)
77. J. Palguta, G. Schubert, B. J. Travis, Fluid flow and chemical alteration in carbonaceous chondrite parent bodies. *Earth Planet. Sci. Lett.* **296**, 235–243 (2010). doi: [10.1016/j.epsl.2010.05.003](https://doi.org/10.1016/j.epsl.2010.05.003)
78. P. A. Bland, B. J. Travis, Giant convecting mud balls of the early solar system. *Sci. Adv.* **3**, e1602514 (2017). doi: [10.1126/sciadv.1602514](https://doi.org/10.1126/sciadv.1602514); pmid: [28740862](https://pubmed.ncbi.nlm.nih.gov/28740862/)
79. Y.-J. Lee, J. W. Morse, Calcite precipitation in synthetic veins: Implications for the time and fluid volume necessary for vein filling. *Chem. Geol.* **156**, 151–170 (1999). doi: [10.1016/S0009-2541\(98\)00183-1](https://doi.org/10.1016/S0009-2541(98)00183-1)
80. M. A. Tyra, J. Farquhar, Y. Guan, L. A. Leshin, An oxygen isotope dichotomy in CM2 chondritic carbonates—A SIMS approach. *Geochim. Cosmochim. Acta* **77**, 383–395 (2012). doi: [10.1016/j.gca.2011.10.003](https://doi.org/10.1016/j.gca.2011.10.003)
81. L. Wilson, K. Keil, S. J. Love, The internal structures and densities of asteroids. *Meteorit. Planet. Sci.* **34**, 479–483 (1999). doi: [10.1111/j.1945-5100.1999.tb01355.x](https://doi.org/10.1111/j.1945-5100.1999.tb01355.x)
82. P. Hoppe, D. MacDougall, G. W. Lugmair, High spatial resolution ion microprobe measurements refine chronology of carbonate formation in Orgueil. *Meteorit. Planet. Sci.* **42**, 1309–1320 (2007). doi: [10.1111/j.1945-5100.2007.tb00576.x](https://doi.org/10.1111/j.1945-5100.2007.tb00576.x)
83. M. Telus, C. M. O. Alexander, E. H. Hauri, J. Wang, Calcite and dolomite formation in the CM parent body: Insight from *in situ* C and O isotope analyses. *Geochim. Cosmochim. Acta* **260**, 275–291 (2019). doi: [10.1016/j.gca.2019.06.012](https://doi.org/10.1016/j.gca.2019.06.012)
84. W. Fujiya, N. Sugiura, H. Hotta, K. Ichimura, Y. Sano, Evidence for the late formation of hydrous asteroids from young meteoritic carbonates. *Nat. Commun.* **3**, 627 (2012). doi: [10.1038/ncomms1635](https://doi.org/10.1038/ncomms1635); pmid: [22252551](https://pubmed.ncbi.nlm.nih.gov/22252551/)
85. C. E. Jilly *et al.*, ⁵³Mn–⁵³Cr dating of aqueously formed carbonates in the CM2 lithology of the Sutter's Mill carbonaceous chondrite. *Meteorit. Planet. Sci.* **49**, 2104–2117 (2014). doi: [10.1111/maps.12305](https://doi.org/10.1111/maps.12305)
86. E. Tonui *et al.*, Petrographic, chemical and spectroscopic evidence for thermal metamorphism in carbonaceous chondrites I: CI and CM chondrites. *Geochim. Cosmochim. Acta* **126**, 284–306 (2014). doi: [10.1016/j.gca.2013.10.053](https://doi.org/10.1016/j.gca.2013.10.053)
87. M. E. Zolensky, N. M. Abreu, M. A. Velbel, A. Rubin, N. Chaumard, T. Noguchi, T. Michikami, in *Primitive Meteorites and Asteroids* (Elsevier, 2018). <https://linkinghub.elsevier.com/retrieve/pii/B9780128133255000021>, pp. 59–204.
88. T. Nakamura, T. Noguchi, M. E. Zolensky, M. Tanaka, Mineralogy and noble-gas signatures of the carbonate-rich lithology of the Tagish Lake carbonaceous chondrite: Evidence for an accretionary breccia. *Earth Planet. Sci. Lett.* **207**, 83–101 (2003). doi: [10.1016/S0012-821X\(02\)00127-5](https://doi.org/10.1016/S0012-821X(02)00127-5)
89. P. G. Brown *et al.*, The fall, recovery, orbit, and composition of the Tagish Lake meteorite: A new type of carbonaceous chondrite. *Science* **290**, 320–325 (2000). doi: [10.1126/science.290.5490.320](https://doi.org/10.1126/science.290.5490.320); pmid: [11030647](https://pubmed.ncbi.nlm.nih.gov/11030647/)
90. C. Lantz *et al.*, Ion irradiation of carbonaceous chondrites: A new view of space weathering on primitive asteroids. *Icarus* **285**, 43–57 (2017). doi: [10.1016/j.icarus.2016.12.019](https://doi.org/10.1016/j.icarus.2016.12.019)
91. M. S. Thompson, M. J. Loeffler, R. V. Morris, L. P. Keller, R. Christofferson, Spectral and chemical effects of simulated space weathering of the Murchison CM2 carbonaceous chondrite. *Icarus* **319**, 499–511 (2019). doi: [10.1016/j.icarus.2018.09.022](https://doi.org/10.1016/j.icarus.2018.09.022)
92. D. S. Lauretta *et al.*, Episodes of particle ejection from the surface of the active asteroid (101955) Bennu. *Science* **366**, eaay3544 (2019). doi: [10.1126/science.aay3544](https://doi.org/10.1126/science.aay3544); pmid: [31806784](https://pubmed.ncbi.nlm.nih.gov/31806784/)
93. E. B. Bierhaus *et al.*, The OSIRIS-REX Spacecraft and the Touch-and-Go Sample Acquisition Mechanism (TAGSAM). *Space Sci. Rev.* **214**, 107 (2018). doi: [10.1007/s11214-018-0521-6](https://doi.org/10.1007/s11214-018-0521-6)
94. E. K. Tonui, M. E. Zolensky, M. E. Lipschutz, M.-S. Wang, T. Nakamura, Yamato 86029: Aqueously altered and thermally metamorphosed CI-like chondrite with unusual textures. *Meteorit. Planet. Sci.* **38**, 269–292 (2003). doi: [10.1111/j.1945-5100.2003.tb00264.x](https://doi.org/10.1111/j.1945-5100.2003.tb00264.x)
95. W. Nozaki, T. Nakamura, T. Noguchi, Bulk mineralogical changes of hydrous micrometeorites during heating in the upper atmosphere at temperatures below 1000 °C. *Meteorit. Planet. Sci.* **41**, 1095–1114 (2006). doi: [10.1111/j.1945-5100.2006.tb00507.x](https://doi.org/10.1111/j.1945-5100.2006.tb00507.x)
96. C. M. O. Alexander, G. D. Cody, B. T. De Gregorio, L. R. Nittler, R. M. Stroud, The nature, origin and modification of insoluble organic matter in chondrites, the major source of Earth's C and N. *Chem. Erde Geochem.* **77**, 227–256 (2017). doi: [10.1016/j.chemer.2017.01.007](https://doi.org/10.1016/j.chemer.2017.01.007); pmid: [31007270](https://pubmed.ncbi.nlm.nih.gov/31007270/)
97. Y. Kebukawa, S. Nakashima, M. E. Zolensky, Kinetics of organic matter degradation in the Murchison meteorite for the evaluation of parent-body temperature history. *Meteorit. Planet. Sci.* **45**, 99–113 (2010). doi: [10.1111/j.1945-5100.2009.01008.x](https://doi.org/10.1111/j.1945-5100.2009.01008.x)
98. H. Y. McSween Jr. *et al.*, Carbonaceous chondrites as analogs for the composition and alteration of Ceres. *Meteorit. Planet. Sci.* **53**, 1793–1804 (2017). doi: [10.1111/maps.12947](https://doi.org/10.1111/maps.12947)
99. H. Hiesinger *et al.*, Cratering on Ceres: Implications for its crust and evolution. *Science* **353**, aaf4759 (2016). doi: [10.1126/science.aaf4759](https://doi.org/10.1126/science.aaf4759); pmid: [27701089](https://pubmed.ncbi.nlm.nih.gov/27701089/)

ACKNOWLEDGMENTS

We are grateful to C. W. V. Wolner for editorial review. We acknowledge the entire OSIRIS-REx Team for making the encounter with Bennu possible. **Funding:** Supported by NASA under Contract NNM10AA11C issued through the New Frontiers Program. **Author contributions:** H.H.K., D.S.L., A.A.S., and V.E.H. contributed to the conceptualization, formal analysis, and methodology of the paper; H.H.K., A.A.S., V.E.H., C.A.B., D.N.D., D.R.G., D.C.R., J.P.E., K.N.B., X.-D.Z., M.G.B., O.S.B., and J.A.S. performed data curation; H.H.K. was primarily responsible for

visualization; K.I., N.P., and E.R.J. contributed methodology and formal analysis and helped with visualization; T.D.G. carried out investigation; D.S.L. and H.L.E. are responsible for project administration; H.H.K., D.S.L., A.A.S. V.E.H., D.N.D., D.R.G., D.C.R., K.N.B., H.C., H.C.C., J.P.D., J.P.E., D.P.G., T.D.G., R.H., K.I., E.R.J., T.J.M., N.P., S.A.S., S.F., B.E.C., X.-D.Z., M.G.D., O.S.B., and J.A.S. contributed to writing the original draft of this paper and contribution of ideas. **Competing interests:** We declare no competing interests. **Data and materials availability:** OVIRS spectral data from the Reconnaissance A phase are available via the Planetary Data System (PDS) at <https://sbn.psi.edu/pds/>

[resource/orex/ovirs.html](https://sbn.psi.edu/pds/resource/orex/ovirs.html); data from the Nightingale site flyby have filenames starting with the dates of data collection, 2019-10-26 to 2019-10-27. OCAMS (PolyCam) images from the Reconnaissance A phase are available on the PDS at <https://sbn.psi.edu/pds/resource/orex/ocams.html>. File names for the data we used are given in table S3. Calcite, siderite, magnesite, and dolomite laboratory spectra (Figs. 1 to 3) were taken from RELAB (www.planetary.brown.edu/rehab/) and the USGS Spectral Library (www.sciencebase.gov/catalog/item/5807a2a2e4b0841e59e3a18d); the spectra identifiers we used are listed in table S1. A catalog of our spectral classifications is provided in Data S1.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6517/eabc3557/suppl/DC1

Materials and Methods

Figs. S1 to S11

Tables S1 to S4

References (100–112)

Data S1

20 April 2020; accepted 24 September 2020

Published online 8 October 2020

10.1126/science.abc3557

RESEARCH

IN SCIENCE JOURNALS

Edited by Michael Funk

ECOLOGICAL GENOMICS

Untangling the genetics of plasticity

The common buckeye butterfly, *Junonia coenia*, exhibits plastic coloration; it has two color morphs, light tan and dark red, that depend on day length and temperature. By selecting for more and less color plasticity, van der Burg *et al.* generated butterfly lines that were used to map the genetic variants that underlie differential coloration. Genome-wide analysis and RNA sequencing identified the genes most likely to be associated with the differences in color plasticity. Inactivation of genes with CRISPR–Cas9 identified three genes that affected the red phenotype, and other techniques identified cis-regulatory, noncoding genomic variants that were correlated with coloration. From these results, the authors were able to model how genetically encoded plasticity and assimilation of the plastic trait likely evolved. —LMZ *Science*, this issue p. 721



The common buckeye butterfly can exhibit seasonal color variation, a complex trait that evolved through genetic assimilation.

ANTIBIOTICS

Triaging antibiotic use

The use of broad-spectrum, second-line antibiotics in treating urinary tract infections (UTIs) is increasing, likely because of the prevalence of antibiotic resistance. Kanjilal *et al.* applied a machine-learning approach calibrated to local hospital electronic health record data to predict the probability of resistance to first- and second-line antibiotic therapies for uncomplicated UTIs. The algorithm then recommended the least broad-spectrum antibiotic to which a given isolate was predicted to be nonresistant. Use of the pipeline reduced both broad-spectrum and ineffective antibiotic prescription for UTIs in the patient cohort relative to clinicians. —CAC *Sci. Transl. Med.* **12**, eaay5067 (2020).

CLIMATE CHANGE

Thought for food

To have any hope of meeting the central goal of the Paris Agreement, which is to limit global warming to 2°C or less, our carbon emissions must be reduced considerably, including

those coming from agriculture. Clark *et al.* show that even if fossil fuel emissions were eliminated immediately, emissions from the global food system alone would make it impossible to limit warming to 1.5°C and difficult even to realize the 2°C target. Thus, major changes in how food is

produced are needed if we want to meet the goals of the Paris Agreement. —HJS

Science, this issue p. 705

BATTERIES

Layering the charge

Layered metal oxides such as lithium cobalt oxide have attracted great attention for rechargeable batteries. In lithium cells, only the octahedral structure forms, but in sodium cells, trigonal prismatic structures are also possible. However, there is a lack of understanding about how to predict and control the formation of each structure. Zhao *et al.* used the simple properties of ions, namely their charge and their radius appropriately weighted by stoichiometry, to determine whether sodium in the interlayers between the



Carbon emissions from the global food system have large climate impacts.

transition metal or other ion-oxide layers remain octahedral rather than switching over to trigonal prismatic coordination. —MSL

Science, this issue p. 708

CORONAVIRUS

Locking down the SARS-CoV-2 spike

Many efforts to develop therapies against severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) are focused on the spike (S) protein trimer that binds to the host receptor. Structures of trimeric S protein show its receptor-binding domain in either an up or a down conformation. Toelzer *et al.* produced SARS-CoV-2 S in insect cells and determined the structure by cryo-electron microscopy. In their dataset, the closed form was predominant and was stabilized by binding linoleic acid, an essential fatty acid. A similar binding pocket appears to be present in previous highly pathogenic coronaviruses, and past studies suggested links between viral infection and fatty acid metabolism. The pocket could be exploited to develop inhibitors that trap S protein in the closed conformation. —VV

Science, this issue p. 725

ECOLOGICAL TRACKING

Ecological “big data”

Human activities are rapidly altering the natural world. Nowhere is this more evident, perhaps, than in the Arctic, yet this region remains one of the most remote and difficult to study. Researchers have increasingly relied on animal tracking data in these regions to understand individual species' responses, but if we want to understand larger-scale change, we need to integrate our understanding across species. Davidson *et al.* introduce an open-source data archive that currently hosts more than 15 million location data points across 96 species and use it to show distinct climate change

responses across species. Such ecological “big data” can lead to a wider understanding of change. —SNV

Science, this issue p. 712

ARCHAEOLOGY

A woman hunter in the early Andes

The ethnographic record of modern hunter-gatherers indicates that men were the primary, almost exclusive, hunters of big game. However, to what extent this pattern existed in prehistory has been challenged on both empirical and theoretical grounds. To examine this question, Haas *et al.* studied the artifacts recovered from a burial of a young woman at a site on the Andean Altiplano dated to more than 8000 years ago. Found with her was a tool kit used in animal processing and numerous projectile points, strongly indicating that she hunted. Additional examples of female hunters at late Pleistocene and early Holocene sites in the Americas were also identified, findings that support a model of relatively undifferentiated subsistence labor among early populations in the Americas. —MSA

Sci. Adv. 10.1126/sciadv.abd0310 (2020).

ORGANIC DEVICES

Low-noise organic photodetectors

Although the high performance and low cost of visible light silicon photodetectors suffice for most applications, their small size and rigid nature are not optimal for applications such as wearable sensors. Organic photodetectors can be flexible and large in area but are often noisy. Fuentes-Hernandez *et al.* found that optimized choices of the semiconductor and electrode materials that improve diode characteristics enable organic photodetectors that can detect low light levels with low noise. —PDS

Science, this issue p. 698

IN OTHER JOURNALS

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



EPIDEMIOLOGY

Famine's intergenerational partner

Famine extends its dangerous reach across generations: Witness the epigenetic effects of the Dutch famine in 1944–1955, which led to diabetes and metabolic disease among the grandchildren of those affected. Sixty years ago (1958–1962), a great famine caused by a collapse of the food system prevailed across China. Chen *et al.* investigated whether the high levels of pulmonary tuberculosis (TB) currently found in Sichuan Province might be a consequence of that famine two generations later. The authors accessed more than 1 million TB cases diagnosed between 2005 and 2018 and identified 12,000 cases that could be attributable to the famine. Malnutrition and accompanying immune suppression are likely to activate latent TB. Subsequently, among the low-birthweight famine cohort (F_1), starvation was associated with greater risk of adult TB emerging and, furthermore, of an increased risk of adult TB emerging in their children (F_2). This effect is most likely mediated by intergenerational damage to immune responses and lung anatomy rather than just exposure to infected family members. The work starkly shows how societal shocks can cut across generations. —CA

Proc. Natl. Acad. Sci. U.S.A. 10.1073/pnas.2008336117 (2020).

Survivors of the Great Chinese Famine of 1958–62 suffered starvation, which has left a legacy of tuberculosis that still prevails today.

NEUROSCIENCE

Change with distance from the soma

Hippocampal neurons receive and integrate synaptic input along their dendritic tree. Inputs located near the cell soma or in the distal dendrite contribute differently to neuronal integration by a variety of mechanisms, including NMDA receptor (NMDAR)–dependent plasticity processes. Using superresolution microscopy, single-nanoparticle imaging, and glutamate uncaging, Ferreira *et al.* investigated the nanoscale organization of NMDARs containing the subunits GluN2A and GluN2B along the dendritic tree. The organization and surface dynamics of GluN2B-NMDARs, but not of GluN2A-NMDARs, changed between proximal and distal clusters, with a gradual increase in receptor local density from proximal to distal dendritic segments. At the proximal dendrite, the nanoscale organization and membrane dynamics of GluN2B-NMDARs were influenced by physical interplay with the protein kinase CaMKII. —PRS

Proc. Natl. Acad. Sci. U.S.A. **117**, 24526 (2020).

NEUROSCIENCE

Mitochondrial anchoring in synapses

ATP has a limited diffusion capacity in axons, so locally anchored presynaptic mitochondria may be used to support high metabolic demand. Only 33% of presynaptic active zones in hippocampi retain mitochondria. Sustained synaptic activity is mainly restricted to these mitochondria-containing synapses during long-term potentiation. Li *et al.* investigated whether increased energy consumption during intensive synaptic activity contributes to presynaptic energy stress and how mitochondria are recruited to synapses. Syntaphilin (SNPH) anchors axonal mitochondria, and deleting *snph* enhances axonal mitochondrial motility in vitro and in vivo. Synaptic activity induces presynaptic energy stress, which

activates the AMPK-PAK signaling cascade, triggering presynaptic mitochondrial recruitment and SNPH-mediated anchoring. Cross-talk between energy sensing and mitochondrial anchoring enables neurons to capture presynaptic mitochondria in response to local energy stress during intensive synaptic activity. —SMH

Nat. Metab. **2**, 1077 (2020).

MOLECULAR DIFFUSION

Mobile aryl groups on graphene

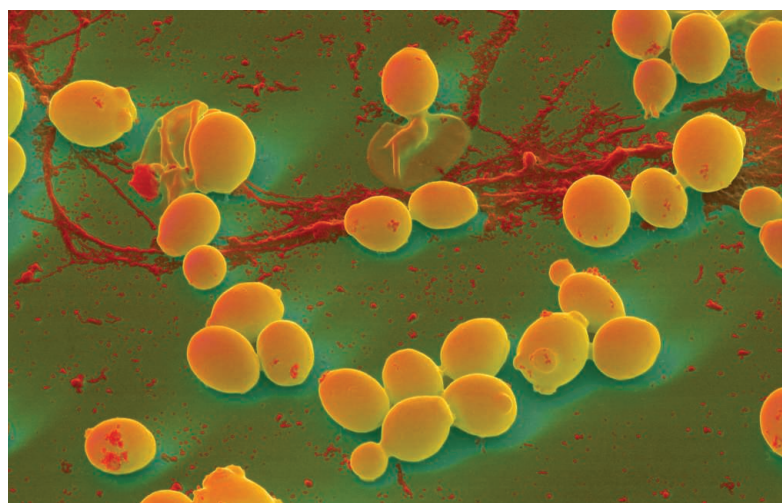
Phenyl groups on aromatic rings can migrate to an adjacent carbon atom under acidic conditions through an ipso arenium ion intermediate. He and Swager show that analogous migration of aryl groups on p-doped graphene surfaces can occur if functional groups weaken the bond to graphene. They patterned graphene with a copper mask to create circular islands of phenyl, 4-trifluoromethylphenyl, and 4-methoxyphenyl groups. Upon oxidative etching of the copper, which leads to p-doping of the graphene, Raman imaging showed that the aryl pattern remained for the first two groups. For 4-methoxyphenyl, migration of these groups several micrometers from their original position occurred through a random walk. —PDS

J. Am. Chem. Soc. **142**, 17876 (2020).

IMMUNOLOGY

Making peace with the beast within

One way that immune cells such as neutrophils and macrophages contribute to host defense is by engulfing microbes. This process, known as phagocytosis, typically entails the formation of an organelle around the microbes called the phagosome, which then fuses with lysosomes to facilitate degradation. However, some pathogens can survive and even thrive within phagosomes. Westman *et al.* investigated the process by which phagosomes sequester and control the rapidly



Candida albicans, shown here, is a yeast-like fungus that can be contained within host phagocytes while immune defenses build.

growing fungus *Candida albicans* after phagocytosis. The macrophage phagosome expands to match the fungus' growth through the calcium-dependent insertion of lysosomes. If lysosomal fusion is inhibited, the phagosomes rupture, leading to NLRP3 inflammasome activation, host cell death, and the enhanced growth and escape of *C. albicans*. However, by tolerating the burgeoning fungus, lysosomal fusion appears to contain infection and give the host precious time to mount its defenses. —STS

Cell Host Microbe **10**, 1016/
j.chom.2020.09.004 (2020).

QUANTUM SYSTEMS

Efficient transfer of entanglement

Processing quantum information across a network relies on the encoding, storage, manipulation, and transfer of quantum states between the nodes along the network chain. However, the information stored in the fragile quantum states can be lost as they propagate along the network, limiting the overall transfer efficiency typically to between 15 and 20% and thereby compromising the scalability of the network. Cao *et al.* developed a method to enhance the efficiency with which the quantum information can be transferred. Using temporally shaped single-photon light pulses and a cold atom storage medium,

their protocol can enhance the storage and retrieval efficiency to ~85%. Further improvement in the transfer efficiency of entanglement between the quantum memories and the information carrier (photons) will advance the development of quantum communication technologies. —ISO

Optica **7**, 1440 (2020).

CHEMICAL STRUCTURE

Computer-aided design of metal borides

Because of their extreme hardness and electronic conductivity, transition-metal borides find wide industrial applications. Expanding their elemental space with new elements can enhance their range of useful properties. Using density functional theory calculations, Dahlqvist *et al.* explored chemical ordering upon M-site alloying of M_2AlB_2 (where M is a transition metal, A is an A-group element, and B is boron) in various symmetries for M from groups 3 to 9. Several previously unknown stable phases were identified, and two of them with in-plane chemical order were successfully synthesized and characterized in agreement with theory. The proposed theoretical approach can serve as a powerful tool to explore the chemical space of metal borides when metal alloying is considered. —YS

J. Am. Chem. Soc. **10**, 1021/
jacs.0c08113 (2020).

ALSO IN *SCIENCE* JOURNALS

Edited by Michael Funk

EPIGENETICS

Linking metabolism to behavior

An emerging theme in studies of animal behavior is that metabolism, as an indicator of the environment, can modulate behavior through epigenetic changes and associated gene expression. In a Perspective, Egervari *et al.* discuss insights emerging from numerous species—including ants, mice, rats, and humans—that the metabolism-epigenetic axis can influence behaviors, including foraging behavior in ants and substance use disorder in rodents and humans. This new paradigm has numerous implications that require further research and could provide a tractable pathway for therapeutic modulation of various disorders. —GKA

Science, this issue p. 660

ASTEROIDS

The complex history of Benu's surface

The near-Earth asteroid (101955) Benu is a carbon-rich body with a rubble pile structure, formed from debris ejected by an impact on a larger parent asteroid. The Origins, Spectral Interpretation, Resource Identification, Security, Regolith Explorer (OSIRIS-REx) spacecraft is designed to collect a sample of Benu's surface and return it to Earth. After arriving at Benu, OSIRIS-REx performed a detailed survey of the asteroid and reconnaissance of potential sites for sample collection. Three papers present results from those mission phases. DellaGiustina *et al.* mapped the optical color and albedo of Benu's surface and established how they relate to boulders and impact craters, finding complex evolution caused by space weathering processes. Simon *et al.* analyzed near-infrared spectra, finding evidence for organic and carbonate materials

that are widely distributed across the surface but are most concentrated on individual boulders. Kaplan *et al.* examined more detailed data collected on the primary sample site, called Nightingale. They identified bright veins with a distinct infrared spectrum in some boulders, which they interpreted as being carbonates formed by aqueous alteration on the parent asteroid. Together, these results constrain Benu's evolution and provide context for the sample collected in October 2020. —KTS

Science, this issue p. 674, p. 675, p. 676

CRIMINAL JUSTICE

“Nudges” and criminal justice

Criminal justice policy in the United States focuses on increasing negative consequences to deter undesired behavior. However, defendants often appear relatively insensitive to these changes in the severity of consequences. Fishbane *et al.* considered a different policy lever: improving the communication of information necessary to adhere to desired behavior (see the Perspective by Kohler-Hausmann). They found that redesigning a criminal summons form to highlight critical information and providing text message reminders increased the likelihood that defendants would show up to their appointed court date, thus eliminating a substantial percentage of arrest warrants for failing to appear in court. In follow-up experiments, the authors found that laypeople, but not experts, believe that such failures to appear are relatively intentional, and this belief reduces their support for interventions aimed at increasing awareness rather than punishment. These findings have implications for policies aimed at improving criminal justice outcomes. —TSR

Science, this issue p. 682;
see also p. 658

METABOLIC GENOMICS

Maximizing elements for your genome

The accumulation of mutations is typically limited by selective parameters. One such parameter is the elements needed to build proteins and molecules to maintain cells. Examining the underlying carbon, oxygen, and nitrogen content for different amino acids, Shenhav *et al.* examined the selective pressure resulting from nutrient limitation (see the Perspective by Polz and Cordero). The authors identified “resource-driven” selection as a purifying selective force associated with environmental nutrient availability, particularly nitrogen, and determined the impact of mutations on the organismal nutritional budget. From this constraint, the authors have proposed that the structure of the genetic code across organisms reflects the mutational impact on elemental resources. —LMZ

Science, this issue p. 683;
see also p. 655

PALEONTOLOGY

Ancient amphibians preserved in amber

Extant amphibians are represented by three fairly simple morphologies: the mostly hopping frogs and toads, the low-crawling salamanders, and the limbless caecilians. Until the early Pleistocene—and for more than 165 million years—there was another group, the albanerpetontids. We know little about this group because amphibian fossils are poorly preserved, and previous specimens from this group are both rare and mostly badly damaged. Daza *et al.* describe a set of fossils preserved in amber showing that this group was unusual both in their habitat use (they may be climbers) and their feeding mode, which appears to have been convergent with the ballistic feeding now

seen in chameleons (see the Perspective by Wake). —SNV

Science, this issue p. 687;
see also p. 654

CORONAVIRUS

Epidemiology in southern India

By August 2020, India had reported several million cases of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), with cases tending to show a younger age distribution than has been reported in higher-income countries. Laxminarayan *et al.* analyzed data from the Indian states of Tamil Nadu and Andhra Pradesh, which have developed rigorous contact tracing and testing systems (see the Perspective by John and Kang). Superspreading predominated, with 5% of infected individuals accounting for 80% of cases. Enhanced transmission risk was apparent among children and young adults, who accounted for one-third of cases. Deaths were concentrated in 50- to 64-year-olds. Incidence did not change in older age groups, possibly because of effective stay-at-home orders and social welfare programs or socioeconomic status. As in other settings, however, mortality rates were associated with older age, comorbidities, and being male. —CA

Science, this issue p. 691;
see also p. 663

CLIMATE CHANGE

The future in the past

A major cause of uncertainties in climate projections is our imprecise knowledge of how much warming should occur as a result of a given increase in the amount of carbon dioxide in the atmosphere. Paleoclimate records have the potential to help us sharpen that understanding because they record

such a wide variety of environmental conditions. Tierney *et al.* review the recent advances in data collection, statistics, and modeling that might help us better understand how rising levels of atmospheric carbon dioxide will affect future climate. —HJS

Science, this issue p. 680

ENHANCER GENOMICS

Enhancer function, from sponges to humans

Identifying the function of enhancers, DNA regions that help to regulate gene expression and evolve rapidly, has been difficult. This area of research has been hampered by the difficulty in identifying functional conservation. Wong *et al.* now show that despite low sequence conservation, enhancer function is strongly conserved through the animal kingdom (see the Perspective by Harmston). Transgenic expression of sponge enhancers in zebrafish and mice demonstrates that these sequences can drive cell type-specific gene expression across species. These results suggest an unexpectedly deep level of conservation of gene regulation across the animal kingdom maintained over the course of metazoan evolution. —LMZ

Science, this issue p. 681;

see also p. 657

TOPOLOGICAL OPTICS

Switching on topology

Ways to control and design topological features in various systems are being studied intensively because the resulting properties tend to be robust against things such as scattering and defects, endowing the system with topological protection. Maczewsky *et al.* now look to another regime in optics to show that optical nonlinearity can induce a topological change in the properties of a photonic lattice. At low excitation power, probe light uniformly leaks into the rest of the lattice, an optically trivial phase. Above

a threshold power, optical nonlinearity induces a topological change in the properties of the photonic lattice, and probe light is confined to propagate along the edge of the structure. These results illustrate a route to dynamically control the propagation of light. —ISO

Science, this issue p. 701

PALEOCLIMATE

Calving cousins

Walczak *et al.* report that increases in Pacific Ocean ventilation and periods of rapid production of icebergs from the Cordilleran Ice Sheet during the last glacial period preceded episodic iceberg discharges into the Atlantic Ocean (see the Perspective by Jaeger and Shevenell). Marine sediments from the Gulf of Alaska show that increases in vertical mixing of the ocean there correspond with intense iceberg calving from the ice sheet that covered much of high-latitude western North America and that these changes occurred before the analogous Heinrich events in the North Atlantic. Thus, these Pacific climate system reorganizations may have been an early part of a cascade of dynamic climate events with global repercussions. —HJS

Science, this issue p. 716;

see also p. 662

TUMOR IMMUNOLOGY

CAR T cells get a STAT tune-up

Adoptive cell therapies using chimeric antigen receptor (CAR) T cells display potent antitumor immunity, but T cell exhaustion can compromise their efficacy. Building on a previous observation that interleukin-7 treatment induced polyfunctional CD4⁺ T cells producing multiple cytokines, Ding *et al.* expressed a constitutively active variant of the signaling protein STAT5 called CASTAT5 in mouse CD4⁺ T cells specific for a tumor antigen. CASTAT5-expressing T cells underwent epigenetic

remodeling, developed polyfunctionality, and promoted CD8⁺ T cell-dependent elimination of tumors. In a mouse model of CAR T cells targeting the B cell antigen CD19, CASTAT5 expression in both CD4⁺ and CD8⁺ T cells resulted in improved tumor elimination. These findings introduce a new T cell-engineering approach for CAR T cells that could yield improved resistance to exhaustion and more robust antitumor activity. —IRW

Sci. Immunol. **5**, eaba5962 (2020).

MICROBIAL SIGNALING

Pathogen rewiring for host adaptation

Clinical isolates of the opportunistic pathogen *Pseudomonas aeruginosa* from cystic fibrosis patients often have mutations that affect the two-component systems—the signaling modules in bacteria. Cao *et al.* found that mutations in *P. aeruginosa* strains adapted to cystic fibrosis patients resulted in disruption of a two-component system and aberrant cross-talk between distinct two-component systems. These changes enabled the strains to switch from an acute to chronic infection state and allowed the activation of a two-component system by glucose, which is abundant in the lungs of cystic fibrosis patients. —AMV

Sci. Signal. **13**, eaaz1529 (2020).

REVIEW SUMMARY

CLIMATE CHANGE

Past climates inform our future

Jessica E. Tierney*, Christopher J. Poulsen, Isabel P. Montañez, Tripti Bhattacharya, Ran Feng, Heather L. Ford, Bärbel Hönlisch, Gordon N. Inglis, Sierra V. Petersen, Navjit Sagoo, Clay R. Tabor, Kaustubh Thirumalai, Jiang Zhu, Natalie J. Burls, Gavin L. Foster, Yves Goddérís, Brian T. Huber, Linda C. Ivany, Sandra Kirtland Turner, Daniel J. Lunt, Jennifer C. McElwain, Benjamin J. W. Mills, Bette L. Otto-Bliesner, Andy Ridgwell, Yi Ge Zhang

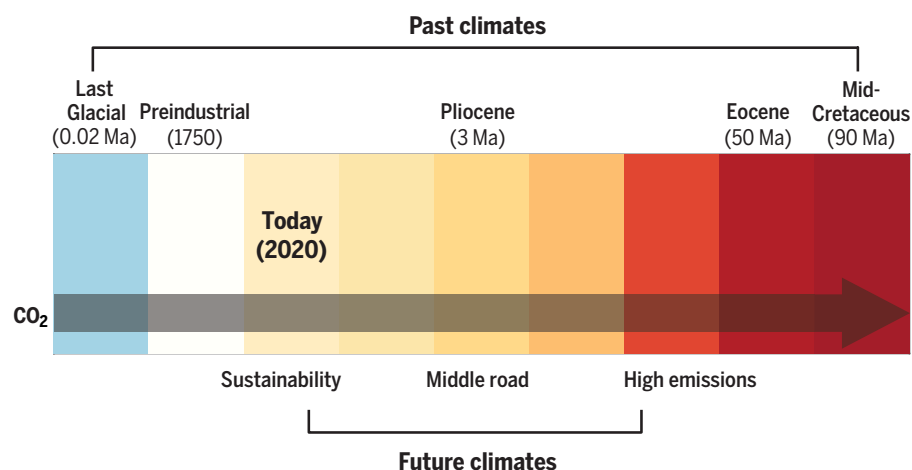
BACKGROUND: Anthropogenic emissions are rapidly altering Earth's climate, pushing it toward a warmer state for which there is no historical precedent. Although no perfect analog exists for such a disruption, Earth's history includes past climate states—"paleoclimates"—that hold lessons for the future of our warming world. These periods in Earth's past span a tremendous range of temperatures, precipitation patterns, cryospheric extent, and biospheric adaptations and are increasingly relevant for improving our understanding of how key elements of the climate system are affected by greenhouse gas levels. The rise of new geochemical and statistical methods, as well as improvements in paleoclimate modeling, allow for formal evaluation of climate models based on paleoclimate data. In particular, given that some of the newest generation of climate models have a high sensitivity to a doubling of atmospheric CO₂, there is a renewed role for paleoclimates in constraining equilibrium climate sensitivity (ECS) and its dependence on climate background state.

ADVANCES: In the past decade, an increasing number of studies have used paleoclimate temperature and CO₂ estimates to infer ECS in the deep past, in both warm and cold climate states. Recent studies support the paradigm that ECS is strongly state-dependent, rising with increased CO₂ concentrations. Simulations of past warm climates such as the Eocene further highlight the role that cloud feedbacks play in contributing to high ECS under increased CO₂ levels. Paleoclimates have provided critical constraints on the assessment of future ice sheet stability and concomitant sea level rise, including the viability of threshold processes like marine ice cliff instability. Beyond global-scale changes, analyses of past changes in the water cycle have advanced our understanding of dynamical drivers of hydroclimate, which is highly relevant for regional climate projections and societal impacts. New and expanding techniques, such as analyses of single shells of foraminifera, are yielding subseasonal climate information that can be used to study how intra- and interannual modes of variability are affected by external

climate forcing. Studies of extraordinary, transient departures in paleoclimate from the background state such as the Paleocene-Eocene Thermal Maximum provide critical context for the current anthropogenic aberration, its impact on the Earth system, and the time scale of recovery.

A number of advances have eroded the "language barrier" between climate model and proxy data, facilitating more direct use of paleoclimate information to constrain model performance. It is increasingly common to incorporate geochemical tracers, such as water isotopes, directly into model simulations, and this practice has vastly improved model-proxy comparisons. The development of new statistical approaches rooted in Bayesian inference has led to a more thorough quantification of paleoclimate data uncertainties. In addition, techniques like data assimilation allow for a formal combination of proxy and model data into hybrid products. Such syntheses provide a full-field view of past climates and can put constraints on climate variables that we have no direct proxies for, such as cloud cover or wind speed.

OUTLOOK: A common concern with using paleoclimate information as model targets is that non-CO₂ forcings, such as aerosols and trace greenhouse gases, are not well known, especially in the distant past. Although evidence thus far suggests that such forcings are secondary to CO₂, future improvements in both geochemical proxies and modeling are on track to tackle this issue. New and rapidly evolving geochemical techniques have the potential to provide improved constraints on the terrestrial biosphere, aerosols, and trace gases; likewise, biogeochemical cycles can now be incorporated into paleoclimate model simulations. Beyond constraining forcings, it is critical that proxy information is transformed into quantitative estimates that account for uncertainties in the proxy system. Statistical tools have already been developed to achieve this, which should make it easier to create robust targets for model evaluation. With this increase in quantification of paleoclimate information, we suggest that modeling centers include simulation of past climates in their evaluation and statement of their model performance. This practice is likely to narrow uncertainties surrounding climate sensitivity, ice sheets, and the water cycle and thus improve future climate projections. ■



Past climates provide context for future climate scenarios. Both past (top) and future (bottom) climates are colored by their estimated change in global mean annual surface temperature relative to preindustrial conditions, ranging from blue (colder) to red (warmer). "Sustainability," "Middle road," and "High emissions" represent the estimated global temperature anomalies at year 2300 from the Shared Socioeconomic Pathways (SSPs) SSP1-2.6, SSP2-4.5, and SSP5-8.5, respectively. In both the past and future cases, warmer climates are associated with increases in CO₂ (indicated by the arrow). Ma, millions of years ago.

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

*Corresponding author. Email: jesst@email.arizona.edu
Cite this article as J. E. Tierney et al., *Science* **370**, eaay3701 (2020). DOI: [10.1126/science.aay3701](https://doi.org/10.1126/science.aay3701)

READ THE FULL ARTICLE AT
<https://doi.org/10.1126/science.aay3701>

REVIEW

CLIMATE CHANGE

Past climates inform our future

Jessica E. Tierney^{1*}, Christopher J. Poulsen², Isabel P. Montañez³, Tripti Bhattacharya⁴, Ran Feng⁵, Heather L. Ford⁶, Bärbel Hönlisch⁷, Gordon N. Inglis⁸, Sierra V. Petersen², Navjit Sagoo⁹, Clay R. Tabor⁵, Kaustubh Thirumalai¹, Jiang Zhu², Natalie J. Burls¹⁰, Gavin L. Foster⁸, Yves Goddérès¹¹, Brian T. Huber¹², Linda C. Ivany⁴, Sandra Kirtland Turner¹³, Daniel J. Lunt¹⁴, Jennifer C. McElwain¹⁵, Benjamin J. W. Mills¹⁶, Bette L. Otto-Bliesner¹⁷, Andy Ridgwell¹³, Yi Ge Zhang¹⁸

As the world warms, there is a profound need to improve projections of climate change. Although the latest Earth system models offer an unprecedented number of features, fundamental uncertainties continue to cloud our view of the future. Past climates provide the only opportunity to observe how the Earth system responds to high carbon dioxide, underlining a fundamental role for paleoclimatology in constraining future climate change. Here, we review the relevancy of paleoclimate information for climate prediction and discuss the prospects for emerging methodologies to further insights gained from past climates. Advances in proxy methods and interpretations pave the way for the use of past climates for model evaluation—a practice that we argue should be widely adopted.

The discipline of paleoclimatology is rooted in the peculiarities of the geological record, which has long hinted that Earth's climate can change in profound ways. In possibly the first paleoclimate study, the 17th-century English physicist Robert Hooke concluded, based on observations of large turtles and ammonites in Jurassic rocks, that conditions in England had once been much warmer than now (1). Since then, paleoclimate studies have revolutionized our view of the climate system (2), documenting both warm and cold worlds much different from the one we inhabit and establishing the link between atmospheric CO₂ and global temperature (Fig. 1).

Although paleoclimatology continues to narrate the history of Earth's climate, it also plays an increasingly central role in understanding future climate change. The study of past

climate has never been more relevant than it is now, as anthropogenic activities increase atmospheric greenhouse gas concentrations and modify the land surface and ocean chemistry at a rate and scale that exceed natural geologic processes. CO₂ levels are higher now than at any point in at least the past 3 million years and, at the current rate of emissions, will exceed concentrations typical of the past 30 million years by 2300 (Fig. 1). In this context, past climates are windows into our future (3)—the geological record is the only observational source of information for how the climate system operates in a state much warmer than the present.

The challenge for paleoclimatology is that there are few direct quantitative records of past climate (e.g., temperature, precipitation). Instead, we make use of “proxies,” surrogates for climate variables that cannot be measured directly. In some cases, the physical occurrence (or absence) of a proxy (like glacial deposits) reveals information about past environmental conditions. More often, geochemical data (such as elemental and stable isotope ratios) stored in fossils, minerals, or organic compounds are used to infer past conditions. The discovery of new proxies, improvements in modeling and analytical techniques, and the increasing number of proxy records are actively expanding the utility of paleoclimate information. These innovations are refining our understanding of how the climate system responds to large changes in atmospheric CO₂ and provide insights into aspects of past climates (such as seasonality and interannual variability) that were heretofore unknowable.

Among the most important contributions that paleoclimatology can make is the evaluation of Earth system models that we rely on for projecting future climate change. The phys-

ical parameterizations in these sophisticated models are often tuned to fit the preindustrial or historical record (4). However, the latter is short in duration and samples a single climate state with a narrow CO₂ range. The performance of climate models under extreme forcing very different from the present is not commonly assessed, even though the models are used to project changes under high-emissions scenarios. When these models are used to simulate past warm climates, they often predict surface temperatures that are too cold and pole-to-equator temperature gradients that are too large (5). However, a new generation of models, alongside developments in proxy techniques and analysis, now provides opportunities to more fully exploit past climates for model evaluation and assessment of key metrics of the climate system.

Past climates inform key processes

Earth's paleoclimate record contains tremendous variability. Over the past 100 million years, the climate gradually transitioned from an ice-free world of exceptional warmth [the mid-Cretaceous, 92 million years ago (Ma); Fig. 1] to the cold ice ages of the past few million years, glacial worlds with kilometers-thick ice caps covering one-fourth of the land surface [such as the Last Glacial Maximum (LGM), 21,000 years ago; Fig. 1]. Between Cretaceous and LGM extremes lie intermediate warm climates such as the early Eocene (53 to 49 Ma) and Pliocene (5.3 to 2.6 Ma) (Fig. 1). This long-term climate transition was far from steady—short-lived hyperthermal events (6) and cold stadials (7) punctuated the slower trends.

Atmospheric CO₂ concentrations generally follow these swings in global temperature (Fig. 1). Geochemical modeling demonstrates that the balance of geological sources (degassing through volcanism) and sinks (weathering and sedimentation) explains the general features of CO₂'s trajectory (8) and establishes causality—high CO₂ leads to high temperatures. The apparent exceptions to this rule, including the end-Cretaceous and early Paleocene (70 to 60 Ma) and the Miocene (23 to 5.3 Ma), are areas of active research. One explanation for the decoupling of CO₂ and temperature is that uncertainties associated with the proxies blur the relationship. Estimation of past CO₂ is challenging. Beyond the ice core record (9), CO₂ information comes from geochemical data, such as isotope ratios of boron and carbon, or paleobotanical indicators such as the density of leaf stomata. All of these proxies require assumptions about the physical, chemical, and biological state of the past that are not completely understood, sometimes leading to misinterpretations of the signal (10). Proxy methodologies and assumptions continue to be refined, and there is some indication that CO₂ at the end of the Cretaceous may have been higher than that shown in Fig. 1 (11). It

¹Department of Geosciences, University of Arizona, Tucson, AZ, USA. ²Department of Earth and Environmental Sciences, University of Michigan, Ann Arbor, MI, USA. ³Department of Earth and Planetary Sciences, University of California, Davis, CA, USA. ⁴Department of Earth and Environmental Sciences, Syracuse University, Syracuse, NY, USA. ⁵Department of Geosciences, University of Connecticut, Storrs, CT, USA. ⁶School of Geography, Queen Mary University of London, London, UK. ⁷Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY, USA. ⁸Department of Earth and Environmental Sciences, Columbia University, Palisades, NY, USA. ⁹School of Ocean and Earth Science, National Oceanography Centre Southampton, University of Southampton, Southampton, UK. ¹⁰Department of Meteorology, University of Stockholm, Stockholm, Sweden. ¹¹Department of Atmospheric, Oceanic, and Earth Sciences, George Mason University, Fairfax, VA, USA. ¹²Centre National de la Recherche Scientifique, Géosciences Environnement Toulouse, Toulouse, France. ¹³Department of Paleobiology, Smithsonian National Museum of Natural History, Washington, DC, USA. ¹⁴Department of Earth Science, University of California, Riverside, Riverside, CA, USA. ¹⁵School of Geographical Sciences, University of Bristol, Bristol, UK. ¹⁶Department of Botany, Trinity College Dublin, Dublin, Ireland. ¹⁷School of Earth and Environment, University of Leeds, Leeds, UK. ¹⁸National Center for Atmospheric Research, Boulder, CO, USA. ¹⁹Department of Oceanography, Texas A&M University, College Station, TX, USA.

*Corresponding author. Email: jess@email.arizona.edu

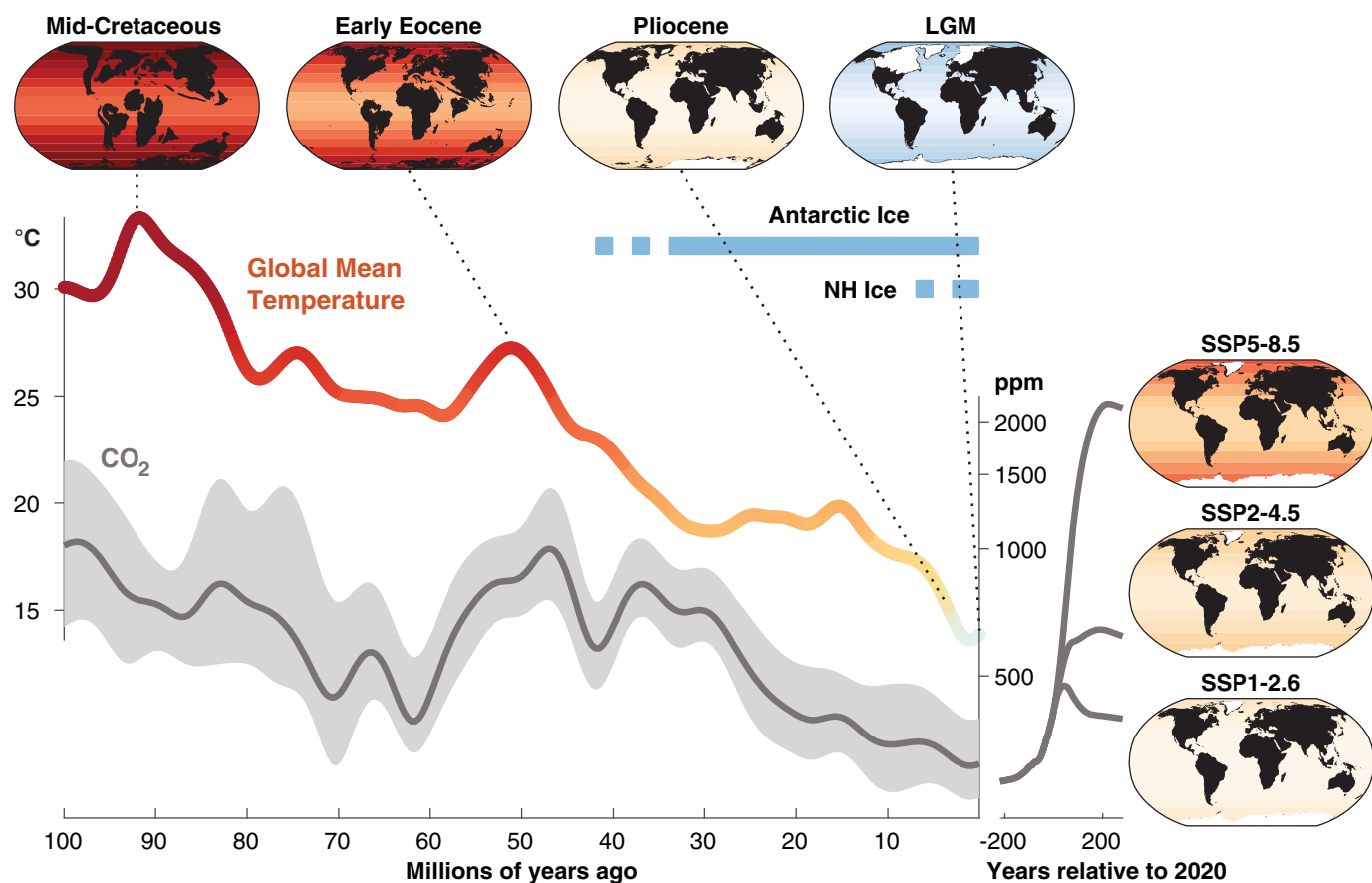


Fig. 1. Paleoclimate context for future climate scenarios. Global mean surface temperature for the past 100 million years is estimated from benthic $\delta^{18}\text{O}$ (2, 106) using the method of (104). CO_2 is estimated from the multiproxy dataset compiled by (105) with additional phytane data from (107) and boron data from (108) and (11). Data with unrealistic values (<150 ppm) are excluded. The CO_2 error envelopes represent 1σ uncertainties. Note the logarithmic scale for CO_2 . Gaussian smoothing was applied

to both the temperature and CO_2 curves to emphasize long-term trends. Temperature colors are scaled relative to preindustrial conditions. The maps show simplified representations of surface temperature. Projected CO_2 concentrations are from the extended SSP scenarios (109). Blue bars indicate when there are well-developed ice sheets (solid lines) and intermittent ice sheets (dashed lines), according to previous syntheses (2). NH, Northern Hemisphere.

is also possible that these discrepancies have another explanation, such as a greater-than-expected role for non- CO_2 forcings and feedbacks. If the paleoclimate record has taught us anything, it is that the more we probe, the more we learn.

Past climate states were profoundly different from today. Their global mean temperatures, latitudinal temperature gradients, polar ice extents, regions of deep-water formation, vegetation types, patterns of precipitation and evaporation, and variability were all different. These differences are invaluable because they provide rich evidence of how climate processes operated across states that span the range of CO_2 concentrations [400 to 2000 parts per million (ppm)] associated with future emissions scenarios [the Shared Socioeconomic Pathways (SSPs); Fig. 1]. Under the sustainable SSP1-2.6 scenario, in which emissions are curtailed and become net negative by the end of the 21st century, CO_2 concentrations

would be stabilized near Pliocene levels (Fig. 1). By contrast, under the fossil-fuel intensive SSP5-8.5 scenario, CO_2 concentrations would approach or even exceed Eocene or mid-Cretaceous levels (Fig. 1). These past warm climates therefore serve as targets against which to measure the increasingly complex generation of climate models that are used for future climate prediction.

Past climates are not perfect analogs for future states—continental configurations are increasingly different with age, and they often represent equilibrium climates as opposed to transient changes associated with rapid greenhouse gas emissions (12). But as benchmarks for climate models, ancient climates need not be perfect analogs. Indeed, the differences are advantageous; they provide true out-of-sample validation for the strength and stability of key feedbacks, large-scale responses of the hydrological cycle, and the most ubiquitous metric of all, climate sensitivity.

Paleoclimate constraints on climate sensitivity

Equilibrium climate sensitivity (ECS) has been widely adopted as a simple metric of how responsive the Earth's climate system is to radiative forcing. It is defined as the change in global near-surface air temperature that results from a sustained doubling in atmospheric CO_2 after the fast-acting (time scales of years to decades) feedback processes (water vapor, clouds, snow) in the Earth system reach equilibrium. The fifth assessment report of the Intergovernmental Panel on Climate Change (IPCC) determined that ECS was likely between 1.5° and 4.5°C , a large range that has remained essentially unchanged for 40 years (13). Because the environmental impacts, socioeconomic implications, and mitigation time scales are very different for a low versus a high ECS (14), narrowing its range has always been a high priority.

That models with either a low or high present-day ECS can match historical observations (15)

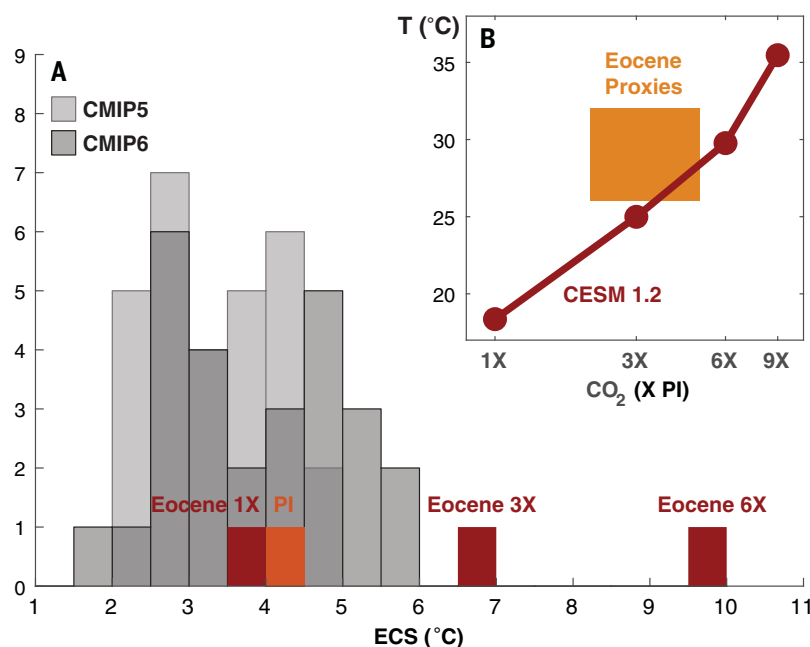


Fig. 2. Constraining ECS through simulation of the early Eocene. (A) ECS in CMIP5 and CMIP6 models [gray bars; (15)] compared with ECS in the CESM1.2 preindustrial (PI, orange bar) and Eocene simulations with 1X, 3X, and 6X preindustrial CO₂ levels (red bars). (B) CO₂ concentrations [times preindustrial level (X PI)] versus global mean temperature according to early Eocene proxies (yellow region) compared with the results from the CESM1.2 Eocene simulations. Proxy CO₂ estimates are a 2 σ range derived from the collection plotted in Fig. 1. See (19) for details of the Eocene climate simulations and proxy global mean temperature estimation.

suggests that preindustrial and industrial climatic changes provide insufficient constraints on this important metric. Furthermore, the emerging view is that ECS is dependent on, and changes with, the background climate state—specifically, it increases in warmer climates (16–19). Past warm climates therefore provide key constraints on the range of plausible ECS values as well as the strength of feedbacks involved. Simulations of the early Eocene provide a salient example. Figure 2 shows a comparison between the ECS of Coupled Model Intercomparison Project (CMIP) phase 5 and phase 6 models (used in the last and the upcoming IPCC assessments) and the ECS of preindustrial and Eocene simulations conducted with the National Center for Atmospheric Research (NCAR) Community Earth System Model (CESM) version 1.2 (19). Doubling CO₂ in an Eocene experiment with preindustrial CO₂ (285 ppm; labeled 1X) yields an ECS similar to the preindustrial experiment and overlaps with the CMIP range (Fig. 2). This indicates that non-CO₂ Eocene boundary conditions, including the position of the continents and the absence of continental ice sheets, do not have a large effect on ECS in CESM1.2. By contrast, raising CO₂ levels increases ECS in the Eocene simulations to values above 6°C (Fig. 2). This increase in ECS with increasing temperature results in accurate simulation of Eocene temperatures at CO₂ concen-

trations that agree with proxy estimates (Fig. 2B). The increased ECS under high CO₂ in CESM1.2 is due to the improved representation of clouds in the CAM5 atmospheric model, which drives a strong low-cloud positive feedback under high CO₂ (19)—a finding in agreement with the emerging recognition that cloud feedbacks are a key component of warm climates (20, 21). That CESM1.2 simulates Eocene proxy temperatures within the bounds of proxy CO₂ estimates provides support for its cloud physics and increases our confidence in the model's state-dependent ECS. The Geophysical Fluid Dynamics Laboratory (GFDL) CM2.1 model can also simulate the large-scale features of Eocene proxy temperatures (22), likewise suggesting that its ECS is reasonable. Conversely, CESM2 (the newest version of the NCAR model) estimates Eocene temperatures that exceed the upper bound of proxy constraints at low CO₂ levels (23), suggesting that its modern ECS of 5.3°C is too high. A little more than a third of the newest-generation CMIP6 models have an ECS higher than 4.5°C (15). From historical observations alone, it is very challenging to assess the plausibility of these higher ECS values. As the Eocene example highlights, warm paleoclimates are key in this respect.

The early Eocene provides an important constraint on model ECS but samples a single

high-CO₂ climate state. Given the dependence of ECS on the background climate state, other past climates are critical to constraining ECS and relevant physics under both lower (e.g., LGM, Pliocene) and higher (e.g., Eocene, Cretaceous) background CO₂ levels. One concern about using past climates as model targets is that the forcings, especially aerosol and non-CO₂ greenhouse gas concentrations, are uncertain and increasingly so in the distant past. Although important, it is worth noting that these forcings are secondary to CO₂ [e.g., (24)] and, for extreme climates like the Eocene and Cretaceous, may largely fall within the climate proxy uncertainties. Moreover, this concern can be mitigated by examining model responses to the potential range of underconstrained forcings and, as is increasingly done, by incorporating biogeochemical cycles and the simulation of aerosol production and transport into the models.

Paleoclimate perspectives on the stability of the cryosphere

Future projections of sea level rise have large uncertainties, mainly owing to unknowns surrounding the stability and threshold behavior of ice sheets (25). The paleoclimate record furnishes true “out-of-sample” tests for understanding the sensitivity of the cryosphere to warming that can lower these uncertainties. The past few years have seen a number of advances on both data and climate modeling fronts to understand past changes in ice sheets and connect these to the future. Advances in the generation and interpretation of proxy indicators of ice sheet size, shape, and extent (26–28) are helping to refine our understanding of cryosphere dynamics in warmer climates. Improvements in modeling the effects of dynamic topography and glacial isostatic adjustment are continually reducing uncertainties associated with estimates of past global sea level (29, 30), providing more accurate benchmarks for model simulations (31).

Paleoclimates also provide critical insights into processes that drive destabilization of ice sheets. Of particular relevance for future projections is assessing the likelihood of marine ice-cliff instability (MICI), a rapid collapse of coastal ice cliffs that follows the disintegration of an ice shelf, which has the potential to contribute to substantial sea level rise by the end of the 21st century (32, 33). The record of sea level change from past warm climates offers a way to test this hypothesis. Recent work has focused on the Pliocene, given that CO₂ concentrations during this time were similar to current anthropogenic levels (Fig. 1). A new reconstruction of global mean sea level during the mid-Pliocene warm period indicates a rise of ~17 m, implying near-to-complete loss of Greenland and the West Antarctic Ice Sheet with some additional contribution from East

Antarctica (34). Although this represents an outstanding loss of ice, MICI is not necessarily needed to explain it (33, 34). However, simulated changes in sea level are highly dependent on each model's treatment of ice sheet stability (35), and paleoclimate investigations of warmer climates, such as the early Pliocene and the Miocene, indicate larger magnitudes of ice loss, thermal expansion, and consequent sea level rise (34, 36). Moving forward, refining our understanding of threshold behavior in ice sheets, and thus improving projections of future sea level rise, will require a synergistic approach that leverages paleoclimate estimates from multiple warm climates alongside solid Earth, ice sheet, and climate modeling (31).

Regional and seasonal information from past climates

Future warming will shift regional and seasonal patterns of rainfall and temperature, with dramatic consequences for human society (37, 38). Regional changes in the land surface (reduced snow cover, melting permafrost, greening, desertification) can further trigger biogeochemical feedbacks that could dampen or amplify initial radiative forcing, with implications for climate sensitivity (39). Unfortunately, climate models disagree about the direction and magnitude of future regional rainfall change (40). Improving future predictions of regional climate requires separating internal variability in the climate system (i.e., interannual-centennial oscillations) from externally forced changes (i.e., from greenhouse gases or aerosols). Regional and seasonal paleoclimate data are critical in this respect, because they provide long, continuous estimates of the natural range of variation, augmenting the relatively short observational record (41, 42).

Subannually resolved paleobiological and sedimentary archives, made more accessible by recent advances in geochemical techniques, allow for the study of seasonal-scale variations in both temperature and hydroclimate. For example, oxygen isotope ($\delta^{18}\text{O}$) measurements of fossil bivalves can be used to gain insights into the drivers of seasonal variability during the Eocene greenhouse climate (43, 44) (Fig. 3A). Because individual planktic foraminifera live for about a month, analyses of single shells yield subannual sea-surface temperature (SST) data from ancient climates (45). This can be leveraged to reveal past changes in key seasonal phenomena such as the El Niño–Southern Oscillation (ENSO) (46) (Fig. 3). Proxy data can even provide records of changes in the frequency or intensity of extreme events like hurricanes (47).

Reconstructions of hydroclimate are considerably more challenging than temperature, because proxy signals tend to be more complex; however, even basic directional information (wetter versus drier) can be used to test

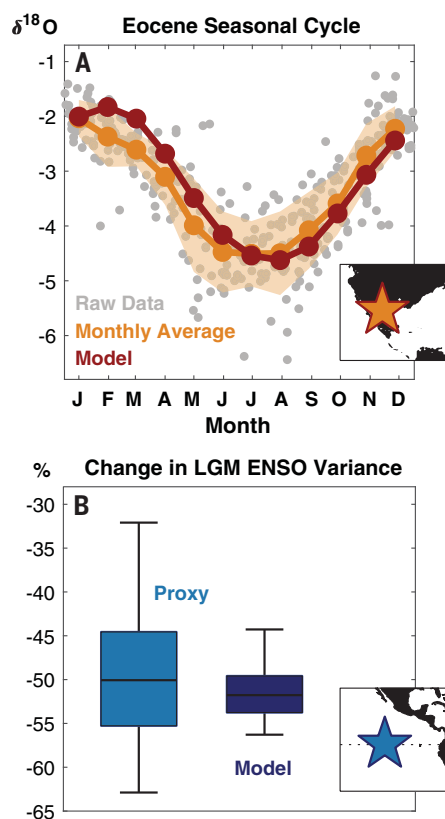


Fig. 3. Examples of seasonal and interannual paleoclimate data and comparison to models.

(A) Seasonally resolved $\delta^{18}\text{O}$ carbonate from the shells of a fossil bivalve, *Venericardia hatcheplatea*, from the early Eocene Hatchetigbee Formation (orange star in inset) (43, 44). Monthly averaged data (orange, with 1σ uncertainty bounds) are compared with predicted $\delta^{18}\text{O}$ -carbonate seasonality at the same grid point from an isotope-enabled Eocene model simulation (19) (red) [using modeled $\delta^{18}\text{O}$ of seawater and SST, and the calibration of (110)]. (B) Mg/Ca measurements of individual planktic foraminifera *Trilobatus sacculifer* from the eastern equatorial Pacific (blue star in inset) provide proxy evidence of a reduction in ENSO variability during the LGM relative to preindustrial conditions (46) (light blue). The magnitude of reduction agrees with simulations using CESM1.2 (dark blue) (111). Error bars represent the 95% confidence interval.

spatial patterns in models [e.g., (48)]. Past warm climates allow us to test the extent to which the thermodynamic “wet-gets-wetter, dry-gets-drier” response broadly holds with warming (49) or if dynamical changes, such as shifts in the Hadley or Walker cells, play more of a key role in the regional water cycle response to changes in surface temperature gradients (48, 50).

Comparisons of proxies and models can also be used to identify the processes that are critical for accurate simulation of regional shifts in the water cycle, where local moisture

and energy budgets exert an important control (51). The processes that drive these budgets—i.e., land surface properties and clouds—must be parameterized in global climate models and are often poorly understood yet have huge consequences for predicted patterns in humidity and rainfall (52–55). Past changes in Earth's boundary conditions offer a much broader set of scenarios where observations can be used to evaluate the performance of parameterization schemes. In particular, paleoclimates spanning the last glacial cycle have helped us better understand the role of land-atmosphere feedbacks in determining hydroclimatic response. Analyses of LGM proxies for SST and water balance in Southeast Asia suggest a direct relationship between convective parameterization and model skill at capturing regional hydroclimate (48, 56). Studies of the mid-Holocene “Green Sahara” highlight the importance of vegetation and dust feedbacks in accurately simulating the response of the west African monsoon to radiative forcing (57, 58). These examples demonstrate the value of hydroclimate proxy-model comparison even if the past climate state is not a direct analog for future warming.

Studies of past warm climates have the potential to provide even more insights into the behavior of regional climate in a warming world. Future model projections broadly simulate a pattern of subtropical drying, whereas the deep tropics and high latitudes get wetter (40). Recently, however, researchers have argued that subtropical drying is transient and might not persist in equilibrium with higher radiative forcing (59, 60). Indeed, several paleoclimatic intervals (61, 62) suggest that a warmer world could feature a different pattern, with wetter conditions in both the subtropics and high latitudes (50). This pattern is especially evident in western North America, where widespread Pliocene lake deposits suggest much wetter conditions (63). This evidence stands in stark contrast to future projections for this region, which overwhelmingly predict drier conditions and more intense droughts (64), and suggests that paleoclimates can provide vital constraints on the response of arid lands to higher CO_2 concentrations.

Climatic aberrations

Among the most important discoveries in paleoclimatology is the occurrence of climatic “aberrations”—extraordinary transient departures from a background climate state. Such events are distinguished by radical changes in temperature, precipitation patterns, and ocean circulation that often leave distinctive marks in the geological record, like the pervasive black shales of the mid-Cretaceous Ocean Anoxic Events (65). An aberration typically occurs in response to a short-lived perturbation to the climate system, such as a sudden release of

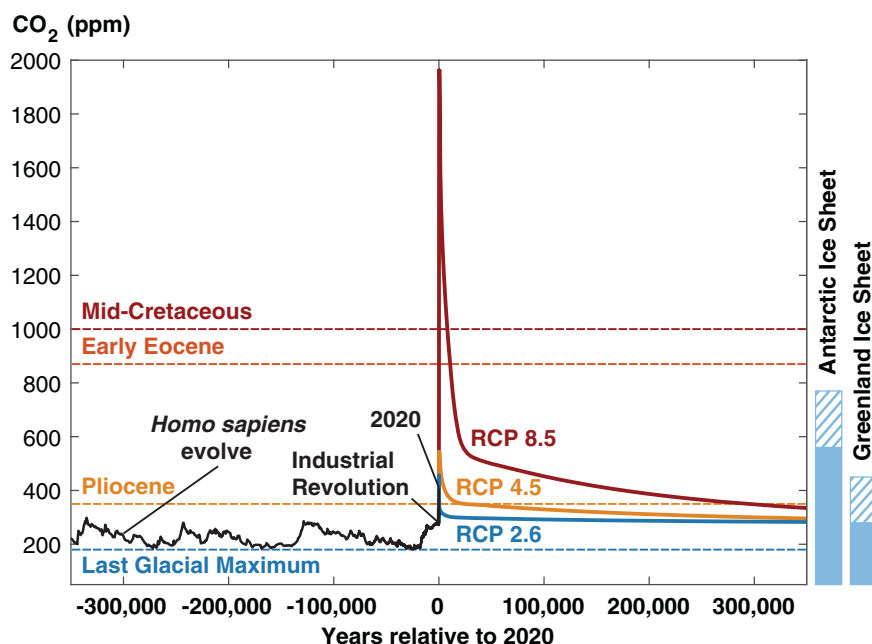


Fig. 4. The anthropogenic climate aberration. The black line shows CO₂ measured in ice cores for the past 350,000 years (9). Solid colored lines show future CO₂ concentrations for the IPCC Fifth Assessment Report RCPs, run out to 350,000 years in the future with the cGENIE model. Dashed lines indicate average CO₂ for key time periods in the geologic past. Bars at the right indicate CO₂ concentrations under which there are well-developed ice sheets (solid areas) and intermittent ice sheets (hatched areas), based on geologic evidence and ice sheet modeling (112).

greenhouse gases (e.g., from volcanoes, methane clathrates, or terrestrial organic deposits). Aberrations need not be “abrupt” in the sense that the rate of climate change must exceed the rate of forcing, and they can potentially last for a long time [for example, the Sturtian Snowball Earth lasted 55 million years (66)]. They are instructive because they provide information on extreme climate states and the ability of the Earth system to recover from such states.

One of the most notable aberrations in the paleoclimate record, the Paleocene-Eocene Thermal Maximum (PETM), may foreshadow future changes that Earth will experience owing to anthropogenic emissions. The PETM, which occurred 56 million years ago, was triggered by rapid emission of greenhouse gases; proxy and model estimates suggest that CO₂ doubled or even tripled from a background state of ~900 ppm (67–69) in less than 5000 years (70, 71). In response, global temperatures spiked by 4° to 6°C (72). The surface ocean rapidly acidified (68, 73) and seafloor carbonates dissolved (74), resulting in dramatic biogeographic range shifts in plankton and the largest extinction in deep-sea calcifying benthic foraminifera ever observed (75). Precipitation patterns changed dramatically, with much more rain falling at the high latitudes (76). It took Earth ~100,000 years to recover from this perturbation (68, 77).

Although the PETM stands out starkly in the geologic record, the rate of CO₂ release was still 4 to 10 times slower than current anthropogenic emissions (71, 78). Indeed, the geological record leaves no doubt that our current rate of global warming, driven by anomalous (anthropogenic) forcing, is an exceptional aberration—the rate and magnitude of change far exceed the typical multi-thousand year variability that preceded it (Fig. 4). In the past 100 million years, CO₂ has ranged from maximum values in the mid-Cretaceous to minimum levels at the LGM (Fig. 1). Going forward, we are on pace to experience an equivalent magnitude of change in atmospheric CO₂ concentrations, in reverse, over a period of time that is more than 10,000 times shorter (Fig. 4). In just over 150 years, we have already raised CO₂ concentrations (now at 410 ppm) to Pliocene levels (Fig. 4). Under a middle-of-the-road emissions scenario such as SSP2-4.5 (or the CMIP5 equivalent, RCP4.5), CO₂ will approach 600 ppm by year 2100, and if we follow the high-emissions SSP5-8.5 (or RCP8.5), CO₂ will rise beyond mid-Cretaceous concentrations (around 1000 ppm) by year 2100 (Fig. 4). In comparison, over the past 800,000 years of geologic history, CO₂ only varied between 180 and 280 ppm (9) (see also Fig. 4).

How long will it take for Earth to neutralize anthropogenic CO₂ and return to prein-

dustrial levels? Earth has the ability to recover from a rapid increase in atmospheric CO₂ concentration—the PETM is a textbook example of this process. Indeed, in every case of past CO₂ perturbation, the Earth system has compensated to avoid a runaway greenhouse or a permanent icehouse. Yet the natural recovery from aberrations takes place on geological, not anthropogenic, time scales (Fig. 4). Some of the processes that remove CO₂ from the atmosphere occur on relatively short (100 to 1000 year) time scales (e.g., ocean uptake), but others take tens to hundreds of thousands of years (e.g., weathering of silicate rocks) (79). Using the intermediate-complexity Earth system model cGENIE, we can estimate how long the recovery process takes under different future forcing scenarios. Under an aggressive mitigation scenario [Representative Concentration Pathway (RCP) 2.6], CO₂ concentrations remain at Pliocene-like concentrations (>350 ppm) through the year 2350, but it still takes hundreds of thousands of years for concentrations to return to preindustrial levels (Fig. 4). Under a middle-of-the-road scenario (RCP 4.5), CO₂ peaks around 550 ppm and remains above Pliocene levels for 30,000 years. Under a worst-case scenario (RCP 8.5), atmospheric CO₂ will remain at mid-Cretaceous (>1000 ppm) concentrations for 5000 years, at Eocene concentrations (~850 ppm) for 10,000 years, and at Pliocene concentrations (>350 ppm) for 300,000 years (Fig. 4). It will be at least 500,000 years, a duration equivalent to 20,000 human generations, before atmospheric CO₂ fully returns to pre-industrial levels. Our planet will recover, but for humans, and the organisms with which we share this planet, the changes in climate will appear to be a permanent state shift.

Bridging the gap between paleoclimate data and models

Climate models provide direct estimates of quantities like temperatures, wind speed, and precipitation. By contrast, paleoclimate information is indirect, filtered through a proxy—a physical, chemical, and/or biological entity that responds to climate—such as foraminifera, algae, or the chemical composition of sediments. Proxies are imperfect recorders of climate; they have inherent uncertainties associated with, for example, biological processes and preservation. Thus, although proxy data can be transformed into climate variables for direct comparison with models using regression, transfer functions, and assumptions, if these structural uncertainties are not accounted for, they can lead to unclear or erroneous interpretations. This creates a “language barrier” between model output and proxy data that has limited the use of paleoclimate information to infer past climate states and evaluate climate models. Three key innovations are now breaking down this barrier, allowing paleoclimate

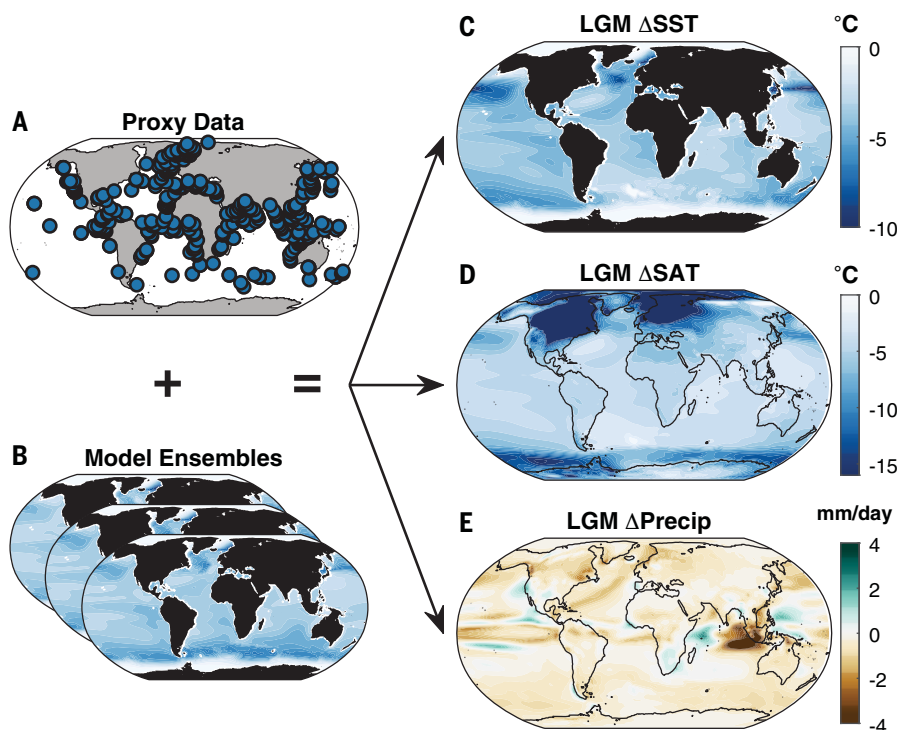


Fig. 5. An example of paleoclimate data assimilation. (A to E) Marine SST proxy data from the LGM and the Late Holocene (A) are combined with an ensemble of model simulations (B) that contain multiple climatic variables. The results [(C) to (E): LGM–Late Holocene differences for SST (Δ SST), surface air temperature (Δ SAT), and mean annual precipitation (Δ Precip)] include all the variables in the model prior, which are influenced by the assimilated SST proxy data. Proxy data, model fields, and assimilated results are from (91).

information to directly constrain model performance: (i) the inclusion of chemical tracers relevant to proxies directly in Earth system models; (ii) the creation of robust proxy system models that explicitly encode processes, uncertainties, and multivariate sensitivities; and (iii) the development of statistical methods to formally combine proxy and model data.

As far as chemical tracers are concerned, the single most important advance has been the increasingly routine incorporation of water isotopes in model simulations. The stable isotopes of water— $\delta^{18}\text{O}$ and hydrogen isotopes (δD)—and their incorporation into natural archives are the foundation of modern paleoclimatology (80). A large number of paleoclimate proxies record water isotopes, including foraminifera, stalagmites, leaf waxes, soil carbonates, and ice cores. Water isotope composition, however, reflects multiple processes, including changes in temperature, moisture source, evaporation, precipitation, and convection. Including water isotopes in models generates simulated isotope fields that are consistent with the model's treatment of these processes, eliminating the need to independently conjecture how these various factors may have influenced the proxy data. This creates an “apples to apples” comparison between proxy information and model output that can be used to evaluate model

performance and diagnose climatic processes [e.g., (81)]. For example, using the water isotope-enabled CESM1.2 (iCESM) (82), it is possible to directly compare carbonate $\delta^{18}\text{O}$ data from Eocene fossil bivalves to model-simulated $\delta^{18}\text{O}$ (43, 44, 83) (Fig. 3A). The model predicts a roughly 3 per mil annual range in carbonate $\delta^{18}\text{O}$, in good agreement with observed proxy data (Fig. 3A). The match with the $\delta^{18}\text{O}$ data builds confidence that the model can correctly simulate climatology in this location and allows us to deconvolve the contribution of SST and $\delta^{18}\text{O}$ of seawater. The site-specific seasonality in SST is 8° to 10°C , whereas the $\delta^{18}\text{O}$ of seawater has a seasonal range of 0.6 to 0.8. This indicates that temperature is primarily responsible for the large seasonal range in carbonate $\delta^{18}\text{O}$ during this greenhouse climate state.

One aspect of paleoclimate information that has traditionally limited its use in model evaluation is an inability to precisely quantify uncertainties surrounding the proxies. However, in the past decade, increasingly detailed proxy system models (84) have been developed to address this issue [e.g., (85–87)]. Many of these use Bayesian inference to quantify uncertainties in the sensitivity of proxies to environmental parameters, which can then be used for probabilistic assessments of past

climate states, model-proxy agreement, and model evaluation (88). These have helped to transform proxy-model comparisons from qualitative statements (“they look similar”) to quantitative statements (“there is a 90% probability that the data and the model agree”).

A final component of the “language barrier” is the fact that proxy data are sparse in both space and time, because they are fundamentally dependent on the presence and preservation of their archives. Yet proxy data are real-world estimates of the “true” climate state. By contrast, climate model information is spatially and temporally continuous and physically self-consistent—but is only a best “guess” at what did or what will happen. One solution to bridge these fundamentally different pieces of information is to formally combine them in a statistical framework and thus leverage their respective strengths. Reduced space methods, commonly used to produce historical reconstructions of climate, can be used to infill missing data and produce maps of paleoclimate states (88, 89). Recently, weather-based data assimilation techniques have been adapted for paleoclimate applications (90). The resulting products are spatially complete reconstructions of multiple climate variables that represent a balance between the proxy information and the physics and covariance structure of the climate model. This allows local paleoclimate proxy information to be used to infer global metrics of climate, such as global mean air temperature (91). It also allows for the recovery of climatic variables that are consistent with the proxy information but for which we have no direct proxies, such as cloud cover, wind patterns, or precipitation (Fig. 5).

In sum, the disintegration of the model-proxy language barrier has narrowed uncertainties in proxy interpretation. Recent studies have been able to use proxy data to infer key climatic processes and evaluate models across multiple time periods, including the LGM (91), the Pliocene (88), and the Eocene (19, 22). This opens the door for explicit use of paleoclimates to assess and improve model physics.

Moving forward

Past climates will continue to provide insights into the range, rate, and dynamics of climate change. Over the past decade, we have witnessed breakthroughs in proxy development and refinement as well as the generation of many new high-resolution marine and terrestrial paleoclimate records. In addition to continued advances, the collection of additional temperature and CO_2 proxy records at higher resolution will be paramount for developing better estimates of climate sensitivity. Future proxy collection efforts should also focus on hydroclimate proxies, given the large spread in model projections (40). These

reconstructions will help us refine our understanding of the response of atmospheric circulation and rainfall to climate change.

On the modeling side, the inclusion of chemical tracers, such as water and carbon isotopes, within many of the newly developed CMIP6 (92) models offers more robust means of data-model comparison. With these new model tools, we anticipate the rapid development and improvement of data-model synthesis products (90, 91) and more focused proxy collection efforts to help reduce model uncertainties. In addition, evaluating CMIP6 models using both the historical and paleoclimate record will result in a more comprehensive and robust approach to understanding the climate system (93). We recommend widespread adoption of this practice, so that model ECS and other emergent properties are constrained by paleoclimate data as well as observations. We suggest that weighting or ranking models that perform well over multiple past climate states is a crucial way to constrain the response of the model to changing background conditions and the validity of simulated climate changes under various emissions scenarios. In general, climate models should be able to accurately simulate multiple extreme paleoclimate states—warm and cold—before being trusted for future climate projection.

Despite promising CMIP6 model advances, maintaining a variety of models with different levels of complexity is important. Not all climate questions require high levels of model complexity, and sometimes complexity is so great that interpretation becomes limited (94). In paleoclimatology, complexity can also lead to prohibitive computational expense. Maintained support for lower-resolution and variable-resolution configurations is vital for better interpreting model results and performing long, transient simulations that can address fundamental aspects of paleoclimatology such as glacial cycles and carbon cycle changes.

Looking ahead, there are many outstanding process-based uncertainties associated with future climate change that paleoclimatology can help constrain. For example, paleobotanical records can inform plant physiological responses to changes in CO₂ (95), which remain highly uncertain (96) but important for quantifying evapotranspirative and surface runoff fluxes. Similarly, past vegetation reconstructions can assess dynamic vegetation models and simulated changes in the hydrologic cycle through time (97). Moreover, additional quantitative reconstructions of hydroclimate, in combination with better constraints on plant physiological functioning in the past, will help refine our understanding of the regional water cycle and its dependence on local energy fluxes and large-scale circulation.

New geochemical techniques will also refine our understanding of the Earth system. Devel-

opment of radiation (98), biogenic aerosol (99), and dust (100) records has the potential to help constrain past aerosol and cloud radiative effects, which are arguably the most uncertain component of Earth system models (101). In addition, new geochemical tracers for methane cycling (102) and upwelling, which is important for N₂O production (103), will provide insights into trace greenhouse gases during past climate states. The combination of these new techniques will allow the paleoclimate community to better quantify biogeochemical feedbacks and climate sensitivity to greenhouse gas forcings across a range of climate states—and ultimately improve climate forecasts for the coming decades to millennia.

The paleoclimate record is the basis for how we understand the potential range and rate of climate change. Past climates represent the only target for climate model predictions at CO₂ concentrations outside of the narrow historical range and, for this reason, are vital tools for evaluating the newest generation of Earth system models. The study of past climates continues to reveal key insights into Earth's response to increased concentrations of greenhouse gases. Innovations in Earth system models, geochemical techniques, and statistical methods further allow for a more direct connection from the past to the future—worlds for which the preindustrial and industrial climate states provide limited guidance. The future of paleoclimatology is to incorporate past climate information formally in model evaluation, so that we can better predict and plan for the impacts of anthropogenic climate change.

REFERENCES AND NOTES

1. R. Hooke, *The Posthumous Works of Robert Hooke... Containing His Cutlerian Lectures, and Other Discourses, Read at the Meetings of the Illustrious Royal Society... Illustrated with Sculptures* (The Royal Society, London, 1705), pp. 342–343.
2. J. Zachos, M. Pagani, L. Sloan, E. Thomas, K. Billups, Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **292**, 686–693 (2001). doi: [10.1126/science.1059412](https://doi.org/10.1126/science.1059412); pmid: [11326091](https://pubmed.ncbi.nlm.nih.gov/11326091/)
3. K. D. Burke *et al.*, Pliocene and Eocene provide best analogs for near-future climates. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 13288–13293 (2018). doi: [10.1073/pnas.1809600115](https://doi.org/10.1073/pnas.1809600115); pmid: [30530685](https://pubmed.ncbi.nlm.nih.gov/30530685/)
4. G. A. Schmidt *et al.*, Practice and philosophy of climate model tuning across six U.S. modeling centers. *Geosci. Model Dev.* **10**, 3207–3223 (2017). doi: [10.5194/gmd-10-3207-2017](https://doi.org/10.5194/gmd-10-3207-2017); pmid: [30595813](https://pubmed.ncbi.nlm.nih.gov/30595813/)
5. M. Huber, R. Caballero, The early Eocene equable climate problem revisited. *Clim. Past* **7**, 603–633 (2011). doi: [10.5194/cp-7-603-2011](https://doi.org/10.5194/cp-7-603-2011)
6. V. Lauretano, J. C. Zachos, L. J. Lourens, Orbitally paced carbon and deep-sea temperature changes at the peak of the early Eocene climatic optimum. *Paleoceanogr. Paleoclimatol.* **33**, 1050–1065 (2018). doi: [10.1029/2018PA003422](https://doi.org/10.1029/2018PA003422)
7. D. A. Hodell, J. E. T. Channell, J. H. Curtis, O. E. Romero, U. Röhl, Onset of “Hudson Strait” Heinrich events in the eastern North Atlantic at the end of the middle Pleistocene transition (1640 ka)? *Paleoceanography* **23**, PA4218 (2008). doi: [10.1029/2008PA001591](https://doi.org/10.1029/2008PA001591)
8. J. C. Walker, P. Hays, J. F. Kasting, A negative feedback mechanism for the long-term stabilization of Earth's surface temperature. *J. Geophys. Res.* **86**, 9776–9782 (1981). doi: [10.1029/JC086iC10p09776](https://doi.org/10.1029/JC086iC10p09776)
9. B. Bereiter *et al.*, Revision of the EPICA Dome C CO₂ record from 800 to 600 kyr before present. *Geophys. Res. Lett.* **42**, 542–549 (2015). doi: [10.1002/2014GL061957](https://doi.org/10.1002/2014GL061957)
10. W. Konrad, D. L. Royer, P. J. Franks, A. Roth-Nebelsick, Quantitative critique of leaf-based paleo-CO₂ proxies: Consequences for their reliability and applicability. *Geol. J.* **10.1002/gj.3807** (2020). doi: [10.1002/gj.3807](https://doi.org/10.1002/gj.3807)
11. M. J. Henehan *et al.*, Rapid ocean acidification and protracted Earth system recovery followed the end-Cretaceous Chicxulub impact. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 22500–22504 (2019). doi: [10.1073/pnas.1905989116](https://doi.org/10.1073/pnas.1905989116); pmid: [31636204](https://pubmed.ncbi.nlm.nih.gov/31636204/)
12. Y. Goddérès, Y. Donnadieu, G. Le Hir, V. Lefebvre, E. Nardin, The role of palaeogeography in the Phanerozoic history of atmospheric CO₂ and climate. *Earth Sci. Rev.* **128**, 122–138 (2014). doi: [10.1016/j.earscirev.2013.11.004](https://doi.org/10.1016/j.earscirev.2013.11.004)
13. J. G. Charney *et al.*, “Carbon dioxide and climate: A scientific assessment” (Consensus Study Report, National Academy of Sciences, The National Academies Press, Washington, DC, 1979).
14. C. Hope, The \$10 trillion value of better information about the transient climate response. *Philos. Trans. A Math. Phys. Eng. Sci.* **373**, 20140429 (2015). doi: [10.1098/rsta.2014.0429](https://doi.org/10.1098/rsta.2014.0429); pmid: [26438286](https://pubmed.ncbi.nlm.nih.gov/26438286/)
15. M. D. Zelinka *et al.*, Causes of higher climate sensitivity in CMIP6 models. *Geophys. Res. Lett.* **47**, e2019GL085782 (2020). doi: [10.1029/2019GL085782](https://doi.org/10.1029/2019GL085782)
16. R. Caballero, M. Huber, State-dependent climate sensitivity in past warm climates and its implications for future climate projections. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 14162–14167 (2013). doi: [10.1073/pnas.1303365110](https://doi.org/10.1073/pnas.1303365110); pmid: [23918397](https://pubmed.ncbi.nlm.nih.gov/23918397/)
17. K. Meraner, T. Mauritsen, A. Voigt, Robust increase in equilibrium climate sensitivity under global warming. *Geophys. Res. Lett.* **40**, 5944–5948 (2013). doi: [10.1002/2013GL058118](https://doi.org/10.1002/2013GL058118)
18. T. Friedrich, A. Timmermann, M. Tigchelaar, O. Elison Timm, A. Ganopolski, Nonlinear climate sensitivity and its implications for future greenhouse warming. *Sci. Adv.* **2**, e1501923 (2016). doi: [10.1126/sciadv.1501923](https://doi.org/10.1126/sciadv.1501923); pmid: [28861462](https://pubmed.ncbi.nlm.nih.gov/28861462/)
19. J. Zhu, C. J. Poulsen, J. E. Tierney, Simulation of Eocene extreme warmth and high climate sensitivity through cloud feedbacks. *Sci. Adv.* **5**, eaax1874 (2019). doi: [10.1126/sciadv.aax1874](https://doi.org/10.1126/sciadv.aax1874); pmid: [31555736](https://pubmed.ncbi.nlm.nih.gov/31555736/)
20. C. S. Bretherton, Insights into low-latitude cloud feedbacks from high-resolution models. *Philos. Trans. A Math. Phys. Eng. Sci.* **373**, 20140415 (2015). doi: [10.1098/rsta.2014.0415](https://doi.org/10.1098/rsta.2014.0415); pmid: [26438280](https://pubmed.ncbi.nlm.nih.gov/26438280/)
21. T. Schneider, C. M. Kaul, K. G. Pressel, Possible climate transitions from breakup of stromatolite decks under greenhouse warming. *Nat. Geosci.* **12**, 163–167 (2019). doi: [10.1038/s41561-019-0310-1](https://doi.org/10.1038/s41561-019-0310-1)
22. D. J. Lunt *et al.*, DeepMIP: Model intercomparison of early Eocene climatic optimum (EEOO) large-scale climate features and comparison with proxy data. *Clim. Past Discuss.* **10.5194/cp-2019-149** (2020). doi: [10.5194/cp-2019-149](https://doi.org/10.5194/cp-2019-149)
23. J. Zhu, C. J. Poulsen, B. L. Otto-Bliesner, High climate sensitivity in CMIP6 model not supported by paleoclimate. *Nat. Clim. Chang.* **10**, 378–379 (2020). doi: [10.1038/s41558-020-0764-6](https://doi.org/10.1038/s41558-020-0764-6)
24. D. J. Beerling, A. Fox, D. S. Stevenson, P. J. Valdes, Enhanced chemistry-climate feedbacks in past greenhouse worlds. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 9770–9775 (2011). doi: [10.1073/pnas.1102409108](https://doi.org/10.1073/pnas.1102409108); pmid: [21628580](https://pubmed.ncbi.nlm.nih.gov/21628580/)
25. J. L. Bamber, M. Oppenheimer, R. E. Kopp, W. P. Aspinall, R. M. Cooke, Ice sheet contributions to future sea-level rise from structured expert judgment. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 11195–11200 (2019). doi: [10.1073/pnas.1817205116](https://doi.org/10.1073/pnas.1817205116); pmid: [31110015](https://pubmed.ncbi.nlm.nih.gov/31110015/)
26. M. G. Wise, J. A. Dowdeswell, M. Jakobsson, R. D. Larter, Evidence of marine ice-cliff instability in Pine Island Bay from ice-berg-keel plough marks. *Nature* **550**, 506–510 (2017). doi: [10.1038/nature24458](https://doi.org/10.1038/nature24458); pmid: [29072274](https://pubmed.ncbi.nlm.nih.gov/29072274/)
27. S. P. S. Gulick *et al.*, Initiation and long-term instability of the East Antarctic Ice Sheet. *Nature* **552**, 225–229 (2017). doi: [10.1038/nature25026](https://doi.org/10.1038/nature25026); pmid: [29239353](https://pubmed.ncbi.nlm.nih.gov/29239353/)
28. A. Rovere *et al.*, The analysis of Last Interglacial (MIS 5e) relative sea-level indicators: Reconstructing sea-level in a warmer world. *Earth Sci. Rev.* **159**, 404–427 (2016). doi: [10.1016/j.earscirev.2016.06.006](https://doi.org/10.1016/j.earscirev.2016.06.006)
29. A. Dutton *et al.*, Sea-level rise due to polar ice-sheet mass loss during past warm periods. *Science* **349**, aaa4019 (2015). doi: [10.1126/science.aaa4019](https://doi.org/10.1126/science.aaa4019); pmid: [26160951](https://pubmed.ncbi.nlm.nih.gov/26160951/)

30. J. Austermann *et al.*, The impact of dynamic topography change on antarctic ice sheet stability during the mid-pliocene warm period. *Geology* **43**, 927–930 (2015). doi: [10.1130/G36988.1](https://doi.org/10.1130/G36988.1)
31. B. P. Horton, R. Kopp, A. Dutton, T. Shaw, Geological records of past sea-level changes as constraints for future projections. *PAGES Magazine* **27**, 28–29 (2019).
32. R. M. DeConto, D. Pollard, Contribution of Antarctica to past and future sea-level rise. *Nature* **531**, 591–597 (2016). doi: [10.1038/nature17145](https://doi.org/10.1038/nature17145); pmid: [27029274](https://pubmed.ncbi.nlm.nih.gov/27029274/)
33. T. L. Edwards *et al.*, Revisiting Antarctic ice loss due to marine ice-cliff instability. *Nature* **566**, 58–64 (2019). doi: [10.1038/s41586-019-0901-4](https://doi.org/10.1038/s41586-019-0901-4); pmid: [30728522](https://pubmed.ncbi.nlm.nih.gov/30728522/)
34. O. A. Dumitru *et al.*, Constraints on global mean sea level during Pliocene warmth. *Nature* **574**, 233–236 (2019). doi: [10.1038/s41586-019-1543-2](https://doi.org/10.1038/s41586-019-1543-2); pmid: [31471591](https://pubmed.ncbi.nlm.nih.gov/31471591/)
35. A. M. Dolan, B. de Boer, J. Bernales, D. J. Hill, A. M. Haywood, High climate model dependency of Pliocene Antarctic ice-sheet predictions. *Nat. Commun.* **9**, 2799 (2018). doi: [10.1038/s41467-018-05179-4](https://doi.org/10.1038/s41467-018-05179-4); pmid: [30022077](https://pubmed.ncbi.nlm.nih.gov/30022077/)
36. E. Gasson, R. M. DeConto, D. Pollard, R. H. Levy, Dynamic Antarctic ice sheet during the early to mid-Miocene. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 3459–3464 (2016). doi: [10.1073/pnas.1516130113](https://doi.org/10.1073/pnas.1516130113); pmid: [26903645](https://pubmed.ncbi.nlm.nih.gov/26903645/)
37. A. J. McMichael, R. E. Woodruff, S. Hales, Climate change and human health: Present and future risks. *Lancet* **367**, 859–869 (2006). doi: [10.1016/S0140-6736\(06\)68079-3](https://doi.org/10.1016/S0140-6736(06)68079-3); pmid: [16530580](https://pubmed.ncbi.nlm.nih.gov/16530580/)
38. R. L. Wilby, A review of climate change impacts on the built environment. *Built Environ.* **33**, 31–45 (2007). doi: [10.2148/benv.33.1.31](https://doi.org/10.2148/benv.33.1.31)
39. A. Arnett *et al.*, Terrestrial biogeochemical feedbacks in the climate system. *Nat. Geosci.* **3**, 525–532 (2010). doi: [10.1038/ngeo905](https://doi.org/10.1038/ngeo905)
40. R. Knutti, J. Sedláček, Robustness and uncertainties in the new CMIP5 climate model projections. *Nat. Clim. Chang.* **3**, 369–373 (2013). doi: [10.1038/nclimate1716](https://doi.org/10.1038/nclimate1716)
41. C. Deser, A. Phillips, V. Bourdette, H. Teng, Uncertainty in climate change projections: The role of internal variability. *Clim. Dyn.* **38**, 527–546 (2012). doi: [10.1007/s00382-010-0977-x](https://doi.org/10.1007/s00382-010-0977-x)
42. PAGES Hydro2k Consortium, Comparing proxy and model estimates of hydroclimate variability and change over the Common Era. *Clim. Past* **13**, 1851–1900 (2017). doi: [10.5194/cp-13-1851-2017](https://doi.org/10.5194/cp-13-1851-2017)
43. L. C. Ivany *et al.*, Intra-annual isotopic variation in *Venericardia* bivalves: Implications for early Eocene temperature, seasonality, and salinity on the US Gulf Coast. *J. Sediment. Res.* **74**, 7–19 (2004). doi: [10.1306/052803740007](https://doi.org/10.1306/052803740007)
44. C. R. Keating-Bitonti, L. C. Ivany, H. P. Affek, P. Douglas, S. D. Samson, Warm, not super-hot, temperatures in the early Eocene subtropics. *Geology* **39**, 771–774 (2011). doi: [10.1130/G32054.1](https://doi.org/10.1130/G32054.1)
45. K. Thirumalai, P. N. DiNezio, J. E. Tierney, M. Puy, M. Mohtadi, An El Niño mode in the glacial Indian Ocean? *Paleoceanogr. Paleoclimatol.* **34**, 1316–1327 (2019). doi: [10.1029/2019PA003669](https://doi.org/10.1029/2019PA003669)
46. H. L. Ford, A. C. Ravelo, P. J. Polissar, Reduced El Niño–Southern Oscillation during the Last Glacial Maximum. *Science* **347**, 255–258 (2015). doi: [10.1126/science.1258437](https://doi.org/10.1126/science.1258437); pmid: [25593181](https://pubmed.ncbi.nlm.nih.gov/25593181/)
47. A. Frappier, T. Knutson, K.-B. Liu, K. Emanuel, Perspective: Coordinating paleoclimate research on tropical cyclones with hurricane-climate theory and modelling. *Tellus A Dyn. Meteorol. Oceanogr.* **59**, 529–537 (2007). doi: [10.1111/j.1600-0870.2007.00250.x](https://doi.org/10.1111/j.1600-0870.2007.00250.x)
48. P. N. DiNezio, J. E. Tierney, The effect of sea level on glacial Indo-Pacific climate. *Nat. Geosci.* **6**, 485–491 (2013). doi: [10.1038/ngeo1823](https://doi.org/10.1038/ngeo1823)
49. I. M. Held, B. J. Soden, Robust responses of the hydrological cycle to global warming. *J. Clim.* **19**, 5686–5699 (2006). doi: [10.1175/JCLI3990.1](https://doi.org/10.1175/JCLI3990.1)
50. N. J. Burls, A. V. Fedorov, Wetter subtropics in a warmer world: Contrasting past and future hydrological cycles. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 12888–12893 (2017). doi: [10.1073/pnas.1703421114](https://doi.org/10.1073/pnas.1703421114); pmid: [29158397](https://pubmed.ncbi.nlm.nih.gov/29158397/)
51. C. Muller, P. O’Gorman, An energetic perspective on the regional response of precipitation to climate change. *Nat. Clim. Chang.* **1**, 266–271 (2011). doi: [10.1038/nclimate1169](https://doi.org/10.1038/nclimate1169)
52. M. P. Byrne, P. A. O’Gorman, The response of precipitation minus evapotranspiration to climate warming: Why the “wet-get-wetter, dry-get-drier” scaling does not hold over land. *J. Clim.* **28**, 8078–8092 (2015). doi: [10.1175/JCLI-D-15-0369.1](https://doi.org/10.1175/JCLI-D-15-0369.1)
53. M. P. Byrne, P. A. O’Gorman, Trends in continental temperature and humidity directly linked to ocean warming. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 4863–4868 (2018). doi: [10.1073/pnas.1722312115](https://doi.org/10.1073/pnas.1722312115); pmid: [29686095](https://pubmed.ncbi.nlm.nih.gov/29686095/)
54. E. Erfani, N. J. Burls, The strength of low-cloud feedbacks and tropical climate: A CESM sensitivity study. *J. Clim.* **32**, 2497–2516 (2019). doi: [10.1175/JCLI-D-18-0551.1](https://doi.org/10.1175/JCLI-D-18-0551.1)
55. S. Bony *et al.*, Clouds, circulation and climate sensitivity. *Nat. Geosci.* **8**, 261–268 (2015). doi: [10.1038/ngeo2398](https://doi.org/10.1038/ngeo2398)
56. P. N. DiNezio *et al.*, Glacial changes in tropical climate amplified by the Indian Ocean. *Sci. Adv.* **4**, eaat9658 (2018). doi: [10.1126/sciadv.aat9658](https://doi.org/10.1126/sciadv.aat9658); pmid: [30547084](https://pubmed.ncbi.nlm.nih.gov/30547084/)
57. J. E. Tierney, F. S. Pausata, P. B. deMenocal, Rainfall regimes of the Green Sahara. *Sci. Adv.* **3**, e1601503 (2017). doi: [10.1126/sciadv.1601503](https://doi.org/10.1126/sciadv.1601503); pmid: [28116352](https://pubmed.ncbi.nlm.nih.gov/28116352/)
58. W. R. Boos, R. L. Korty, Regional energy budget control of the intertropical convergence zone and application to mid-holocene rainfall. *Nat. Geosci.* **9**, 892–897 (2016). doi: [10.1038/ngeo2833](https://doi.org/10.1038/ngeo2833)
59. P. Ceppi, G. Zappa, T. G. Shepherd, J. M. Gregory, Fast and slow components of the extratropical atmospheric circulation response to CO₂ forcing. *J. Clim.* **31**, 1091–1105 (2018). doi: [10.1175/JCLI-D-17-0323.1](https://doi.org/10.1175/JCLI-D-17-0323.1)
60. J. K. Sniderman *et al.*, Southern Hemisphere subtropical drying as a transient response to warming. *Nat. Clim. Chang.* **9**, 232–236 (2019). doi: [10.1038/s41558-019-0397-9](https://doi.org/10.1038/s41558-019-0397-9)
61. R. Feng, C. J. Poulsen, M. Werner, Tropical circulation intensification and tectonic extension recorded by Neogene terrestrial $\delta^{18}\text{O}$ records of the western United States. *Geology* **44**, 971–974 (2016). doi: [10.1130/G38212.1](https://doi.org/10.1130/G38212.1)
62. B. Carrapa, M. Clementz, R. Feng, Ecological and hydroclimate responses to strengthening of the Hadley circulation in South America during the Late Miocene cooling. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 9747–9752 (2019). doi: [10.1073/pnas.1810721116](https://doi.org/10.1073/pnas.1810721116); pmid: [31036635](https://pubmed.ncbi.nlm.nih.gov/31036635/)
63. D. E. Ibarra *et al.*, Warm and cold wet states in the western United States during the Pliocene–Pleistocene. *Geology* **46**, 355–358 (2018). doi: [10.1130/G39962.1](https://doi.org/10.1130/G39962.1)
64. B. I. Cook, T. R. Ault, J. E. Smerdon, Unprecedented 21st century drought risk in the American Southwest and Central Plains. *Sci. Adv.* **1**, e1400082 (2015). doi: [10.1126/sciadv.1400082](https://doi.org/10.1126/sciadv.1400082); pmid: [26601131](https://pubmed.ncbi.nlm.nih.gov/26601131/)
65. H. C. Jenkyns, Geochemistry of oceanic anoxic events. *Geochem. Geophys. Geosyst.* **11**, Q03004 (2010). doi: [10.1029/2009GC002788](https://doi.org/10.1029/2009GC002788)
66. A. D. Rooney *et al.*, Re-Os geochronology and coupled Os–Sr isotope constraints on the Sturtian snowball Earth. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 51–56 (2014). doi: [10.1073/pnas.1317266110](https://doi.org/10.1073/pnas.1317266110); pmid: [24344274](https://pubmed.ncbi.nlm.nih.gov/24344274/)
67. R. E. Zeebe, J. C. Zachos, G. R. Dickens, Carbon dioxide forcing alone insufficient to explain Palaeocene–Eocene Thermal Maximum warming. *Nat. Geosci.* **2**, 576–580 (2009). doi: [10.1038/ngeo578](https://doi.org/10.1038/ngeo578)
68. M. Gutjahr *et al.*, Very large release of mostly volcanic carbon during the Palaeocene–Eocene Thermal Maximum. *Nature* **548**, 573–577 (2017). doi: [10.1038/nature23646](https://doi.org/10.1038/nature23646); pmid: [28858305](https://pubmed.ncbi.nlm.nih.gov/28858305/)
69. Y. Cui, B. A. Schubert, Atmospheric pCO₂ reconstructed across five early Eocene global warming events. *Earth Planet. Sci. Lett.* **478**, 225–233 (2017). doi: [10.1016/j.epsl.2017.08.038](https://doi.org/10.1016/j.epsl.2017.08.038)
70. S. Kirtland Turner, P. M. Hull, L. R. Kump, A. Ridgwell, A probabilistic assessment of the rapidity of PETM onset. *Nat. Commun.* **8**, 353 (2017). doi: [10.1038/s41467-017-00292-2](https://doi.org/10.1038/s41467-017-00292-2); pmid: [28842564](https://pubmed.ncbi.nlm.nih.gov/28842564/)
71. S. Kirtland Turner, Constraints on the onset duration of the Paleocene–Eocene Thermal Maximum. *Philos. Trans. A Math. Phys. Eng. Sci.* **376**, 20170082 (2018). doi: [10.1098/rsta.2017.0082](https://doi.org/10.1098/rsta.2017.0082); pmid: [30177565](https://pubmed.ncbi.nlm.nih.gov/30177565/)
72. G. N. Inglis *et al.*, Global mean surface temperature and climate sensitivity of the EECO, PETM and latest Paleocene. *Clim. Past Discuss.* **10**, 5194/cp-2019-167 (2020). doi: [10.5194/cp-2019-167](https://doi.org/10.5194/cp-2019-167)
73. D. E. Penman, B. Hönisch, R. E. Zeebe, E. Thomas, J. C. Zachos, Rapid and sustained surface ocean acidification during the Paleocene–Eocene Thermal Maximum. *Paleoceanography* **29**, 357–369 (2014). doi: [10.1002/2014PA002621](https://doi.org/10.1002/2014PA002621)
74. J. C. Zachos *et al.*, Rapid acidification of the ocean during the Paleocene–Eocene thermal maximum. *Science* **308**, 1611–1615 (2005). doi: [10.1126/science.1109004](https://doi.org/10.1126/science.1109004); pmid: [15947184](https://pubmed.ncbi.nlm.nih.gov/15947184/)
75. E. Thomas, N. J. Shackleton, The Paleocene–Eocene benthic foraminiferal extinction and stable isotope anomalies. *Geol. Soc. London Spec. Publ.* **101**, 401–441 (1996). doi: [10.1144/GSL.SP.1996.101.01.20](https://doi.org/10.1144/GSL.SP.1996.101.01.20)
76. M. J. Carmichael *et al.*, Hydrological and associated biogeochemical consequences of rapid global warming during the Paleocene–Eocene Thermal Maximum. *Global Planet. Change* **157**, 114–138 (2017). doi: [10.1016/j.gloplacha.2017.07.014](https://doi.org/10.1016/j.gloplacha.2017.07.014)
77. G. J. Bowen, J. C. Zachos, Rapid carbon sequestration at the termination of the Palaeocene–Eocene Thermal Maximum. *Nat. Geosci.* **3**, 866–869 (2010). doi: [10.1038/ngeo1014](https://doi.org/10.1038/ngeo1014)
78. R. E. Zeebe, A. Ridgwell, J. C. Zachos, Anthropogenic carbon release rate unprecedented during the past 66 million years. *Nat. Geosci.* **9**, 325–329 (2016). doi: [10.1038/ngeo2681](https://doi.org/10.1038/ngeo2681)
79. D. Archer, H. Khesghi, E. Maier-Reimer, Multiple timescales for neutralization of fossil fuel CO₂. *Geophys. Res. Lett.* **24**, 405–408 (1997). doi: [10.1029/97GL00168](https://doi.org/10.1029/97GL00168)
80. C. Emiliani, Pleistocene temperatures. *J. Geol.* **63**, 538–578 (1955). doi: [10.1086/626295](https://doi.org/10.1086/626295)
81. J. Zhou, C. Poulsen, D. Pollard, T. White, Simulation of modern and middle Cretaceous marine $\delta^{18}\text{O}$ with an ocean-atmosphere general circulation model. *Paleoceanography* **23**, PA3223 (2008). doi: [10.1029/2008PA001596](https://doi.org/10.1029/2008PA001596)
82. E. Brady *et al.*, The connected isotopic water cycle in the Community Earth System Model version 1. *J. Adv. Model. Earth Syst.* **11**, 2547–2566 (2019). doi: [10.1029/2019MS001663](https://doi.org/10.1029/2019MS001663)
83. J. Zhu *et al.*, Simulation of early Eocene water isotopes using an Earth system model and its implication for past climate reconstruction. *Earth Planet. Sci. Lett.* **537**, 116164 (2020). doi: [10.1016/j.epsl.2020.116164](https://doi.org/10.1016/j.epsl.2020.116164)
84. M. N. Evans, S. E. Tolwinski-Ward, D. M. Thompson, K. J. Anchukaitis, Applications of proxy system modeling in high resolution paleoclimatology. *Quat. Sci. Rev.* **76**, 16–28 (2013). doi: [10.1016/j.quascirev.2013.05.024](https://doi.org/10.1016/j.quascirev.2013.05.024)
85. S. Tolwinski-Ward, K. Anchukaitis, M. Evans, Bayesian parameter estimation and interpretation for an intermediate model of tree-ring width. *Clim. Past* **9**, 1481–1493 (2013). doi: [10.5194/cp-9-1481-2013](https://doi.org/10.5194/cp-9-1481-2013)
86. C. N. Jex, S. Phipps, A. Baker, C. Bradley, Reducing uncertainty in the climatic interpretations of speleothem $\delta^{18}\text{O}$. *Geophys. Res. Lett.* **40**, 2259–2264 (2013). doi: [10.1002/grl.50467](https://doi.org/10.1002/grl.50467)
87. J. E. Tierney, M. P. Tingley, BAYSPLINE: A new calibration for the alkenone paleothermometer. *Paleoceanogr. Paleoclimatol.* **33**, 281–301 (2018). doi: [10.1002/2017PA003201](https://doi.org/10.1002/2017PA003201)
88. J. E. Tierney, A. M. Haywood, R. Feng, T. Bhattacharya, B. L. Otto-Bliesner, Pliocene warmth consistent with greenhouse gas forcing. *Geophys. Res. Lett.* **46**, 9136–9144 (2019). doi: [10.1029/2019GL083802](https://doi.org/10.1029/2019GL083802)
89. E. C. Gill, B. Rajagopalan, P. Molnar, T. M. Marchitto, Reduced-dimension reconstruction of the equatorial Pacific SST and zonal wind fields over the past 10,000 years using Mg/Ca and alkenone records. *Paleoceanography* **31**, 928–952 (2016). doi: [10.1002/2016PA002948](https://doi.org/10.1002/2016PA002948)
90. G. J. Hakim *et al.*, The last millennium climate reanalysis project: Framework and first results. *J. Geophys. Res. D Atmospheres* **121**, 6745–6764 (2016). doi: [10.1002/2016JD024751](https://doi.org/10.1002/2016JD024751)
91. J. E. Tierney *et al.*, Glacial cooling and climate sensitivity revisited. *Nature* **584**, 569–573 (2020). doi: [10.1038/s41586-020-2617-x](https://doi.org/10.1038/s41586-020-2617-x); pmid: [32848226](https://pubmed.ncbi.nlm.nih.gov/32848226/)
92. V. Eyring *et al.*, Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geosci. Model Dev.* **9**, 1937–1958 (2016). doi: [10.5194/gmd-9-1937-2016](https://doi.org/10.5194/gmd-9-1937-2016)
93. G. Schmidt *et al.*, Using palaeo-climate comparisons to constrain future projections in CMIP5. *Clim. Past* **10**, 221–250 (2013). doi: [10.5194/cp-10-221-2014](https://doi.org/10.5194/cp-10-221-2014)
94. I. M. Held, The gap between simulation and understanding in climate modeling. *Bull. Am. Meteorol. Soc.* **86**, 1609–1614 (2005). doi: [10.1175/BAMS-86-11-1609](https://doi.org/10.1175/BAMS-86-11-1609)
95. J. D. White *et al.*, A process-based ecosystem model (Paleo-BGC) to simulate the dynamic response of Late Carboniferous plants to elevated O₂ and aridification. *Am. J. Sci.* **320**, 547–598 (2020). doi: [10.2475/09.2020.01](https://doi.org/10.2475/09.2020.01)
96. G. Damour, T. Simonneau, H. Cochard, L. Urban, An overview of models of stomatal conductance at the leaf level. *Plant Cell Environ.* **33**, 1419–1438 (2010). doi: [10.1111/j.1365-3040.2010.02181.x](https://doi.org/10.1111/j.1365-3040.2010.02181.x); pmid: [20545879](https://pubmed.ncbi.nlm.nih.gov/20545879/)

97. S. Hoetzel, L. Dupont, E. Schefuß, F. Rommerskirchen, G. Wefer, The role of fire in Miocene to Pliocene C₄ grassland and ecosystem evolution. *Nat. Geosci.* **6**, 1027–1030 (2013). doi: [10.1038/ngeo1984](https://doi.org/10.1038/ngeo1984)
98. B. H. Lomax *et al.*, Plant spore walls as a record of long-term changes in ultraviolet-B radiation. *Nat. Geosci.* **1**, 592–596 (2008). doi: [10.1038/ngeo278](https://doi.org/10.1038/ngeo278)
99. S. Schüpbach *et al.*, Greenland records of aerosol source and atmospheric lifetime changes from the Eemian to the Holocene. *Nat. Commun.* **9**, 1476 (2018). doi: [10.1038/s41467-018-03924-3](https://doi.org/10.1038/s41467-018-03924-3); pmid: [29662058](https://pubmed.ncbi.nlm.nih.gov/29662058/)
100. S. Kienast, G. Winckler, J. Lippold, S. Albani, N. Mahowald, Tracing dust input to the global ocean using thorium isotopes in marine sediments: Thoromap. *Global Biogeochem. Cycles* **30**, 1526–1541 (2016). doi: [10.1002/2016GB005408](https://doi.org/10.1002/2016GB005408)
101. J. H. Seinfeld *et al.*, Improving our fundamental understanding of the role of aerosol-cloud interactions in the climate system. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 5781–5790 (2016). doi: [10.1073/pnas.1514043113](https://doi.org/10.1073/pnas.1514043113); pmid: [27222566](https://pubmed.ncbi.nlm.nih.gov/27222566/)
102. G. N. Inglis, B. D. A. Naafs, Y. Zheng, J. Schellekens, R. D. Pancost, $\delta^{13}\text{C}$ values of bacterial hopanoids and leaf waxes as tracers for methanotrophy in peatlands. *Geochim. Cosmochim. Acta* **260**, 244–256 (2019). doi: [10.1016/j.gca.2019.06.030](https://doi.org/10.1016/j.gca.2019.06.030)
103. Y. G. Zhang, M. Pagani, J. Henderiks, H. Ren, A long history of equatorial deep-water upwelling in the Pacific Ocean. *Earth Planet. Sci. Lett.* **467**, 1–9 (2017). doi: [10.1016/j.epsl.2017.03.016](https://doi.org/10.1016/j.epsl.2017.03.016)
104. J. Hansen, M. Sato, G. Russell, P. Kharecha, Climate sensitivity, sea level and atmospheric carbon dioxide. *Philos. Trans. A Math. Phys. Eng. Sci.* **371**, 20120294–20120294 (2013). doi: [10.1098/rsta.2012.0294](https://doi.org/10.1098/rsta.2012.0294); pmid: [24043864](https://pubmed.ncbi.nlm.nih.gov/24043864/)
105. G. L. Foster, D. L. Royer, D. J. Lunt, Future climate forcing potentially without precedent in the last 420 million years. *Nat. Commun.* **8**, 14845 (2017). doi: [10.1038/ncomms14845](https://doi.org/10.1038/ncomms14845); pmid: [28375201](https://pubmed.ncbi.nlm.nih.gov/28375201/)
106. O. Friedrich, R. D. Norris, J. Erbacher, Evolution of middle to Late Cretaceous oceans—A 55 m.y. record of Earth's temperature and carbon cycle. *Geology* **40**, 107–110 (2012). doi: [10.1130/G32701.1](https://doi.org/10.1130/G32701.1)
107. C. R. Witkowski, J. W. H. Weijers, B. Blais, S. Schouten, J. S. Sinninghe Damsté, Molecular fossils from phytoplankton reveal secular Pco₂ trend over the Phanerozoic. *Sci. Adv.* **4**, eaat4556 (2018). doi: [10.1126/sciadv.aat4556](https://doi.org/10.1126/sciadv.aat4556); pmid: [30498776](https://pubmed.ncbi.nlm.nih.gov/30498776/)
108. S. M. Soudian *et al.*, Constraining the evolution of Neogene ocean carbonate chemistry using the boron isotope pH proxy. *Earth Planet. Sci. Lett.* **498**, 362–376 (2018). doi: [10.1016/j.epsl.2018.06.017](https://doi.org/10.1016/j.epsl.2018.06.017)
109. M. Meinshausen *et al.*, The SSP greenhouse gas concentrations and their extensions to 2500. *Geosci. Model Dev. Discuss.* **2019**, 1–77 (2019).
110. E. L. Grossman, T.-L. Ku, Oxygen and carbon isotope fractionation in biogenic aragonite: Temperature effects. *Chem. Geol.* **59**, 59–74 (1986). doi: [10.1016/0168-9622\(86\)90057-6](https://doi.org/10.1016/0168-9622(86)90057-6)
111. J. Zhu *et al.*, Reduced ENSO variability at the LGM revealed by an isotope-enabled Earth system model. *Geophys. Res. Lett.* **44**, 6984–6992 (2017). doi: [10.1002/2017GL073406](https://doi.org/10.1002/2017GL073406)
112. R. M. Deconto *et al.*, Thresholds for Cenozoic bipolar glaciation. *Nature* **455**, 652–656 (2008). doi: [10.1038/nature07337](https://doi.org/10.1038/nature07337); pmid: [18833277](https://pubmed.ncbi.nlm.nih.gov/18833277/)

ACKNOWLEDGMENTS

This paper is an outcome of an Aspen Global Change Institute workshop funded by NASA and the Heising-Simons Foundation.

Funding: J.E.T. and C.J.P. acknowledge support from grant no. 2016-015 from the Heising-Simons Foundation. S.K.T. acknowledges support from grant no. 2015-145 from the Heising-Simons Foundation. B.L.O.-B. acknowledges support from the National Center for Atmospheric Research, which is a major facility sponsored by the National Science Foundation under cooperative agreement no. 1852977. **Author contributions:** Authorship of this manuscript is organized into three tiers according to the contributions of each individual author. J.E.T., C.J.P., and I.P.M. (tier I) organized the concept, structure, and writing of the manuscript; contributed to all sections of the text; designed the final figures; and wrote the Review summary. Tier II authors (listed alphabetically following I.P.M.) assumed a leading role in the development of different sections of the text as well as the figure design. Tier III authors (listed alphabetically following J.Z.) contributed to the text and figure design under the leadership of the other tiers. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The figures in this Review use data from previously published works, with a few exceptions. The scaled temperature record in Fig. 1 follows (104) but extends further back in time; likewise, the CO₂ curve in Fig. 1 follows (105) but has been updated with data published since then. The future projections of CO₂ concentrations calculated with cGENIE shown in Fig. 4 were run for this Review. All of these data are available for download from <https://github.com/jesstierney/PastClimates>.

10.1126/science.aay3701

RESEARCH ARTICLE SUMMARY

ENHANCER GENOMICS

Deep conservation of the enhancer regulatory code in animals

Emily S. Wong^{*}, Dawei Zheng[†], Siew Z. Tan[†], Neil L. Bower, Victoria Garside, Gilles Vanwalleghem, Federico Gaiti, Ethan Scott, Benjamin M. Hogan, Kazu Kikuchi, Edwina McGlinn, Mathias Francois^{*†}, Bernard M. Degnan^{*†}

INTRODUCTION: In animals, gene regulatory networks specify cell identity in space and time. Transcription of genes in these networks is modulated by a class of cis-regulatory elements called enhancers that contain short (~10 base pairs) DNA sequence motifs recognized by transcription factors (TFs). In contrast to TFs, whose histories have been largely traced to the origin of the animal kingdom or earlier, the origin and evolution of enhancers have been relatively difficult to discern.

Although not a single enhancer has been shown to be conserved across the animal kingdom, enhancers may be as ancient and conserved as the TFs with which they interact. This inability to identify conserved enhancers is apparently because they evolve faster than both the TFs they interact with and the genes they regulate.

RATIONALE: Putative enhancers in the sponge *Amphimedon queenslandica* had previously been identified on the basis of combinatorial patterns of histone modifications. Here, we sought to determine whether sponges share functionally conserved enhancers with bilaterians.

We primarily focused on deeply conserved metazoan microsyntenic gene pairs. These pairs are thought to be conserved because the cis-regulatory elements that regulate the developmental expression of one gene (the target gene) are located in the other gene (the bystander gene). This proposed regulatory linkage may underlie the maintenance of these microsyntenic gene pairs across 700 million years of independent evolution.

RESULTS: We found that enhancers present in *Amphimedon* microsyntenic regions drive consistent patterns of cell type-specific gene expres-

sion in zebrafish and mouse embryos. Although these sponge enhancers do not share significant sequence identity with vertebrates, they are in microsyntenic regions that are orthologous with microsyntenic regions in other metazoans and have strong histone H3 Lys⁴ methylation (H3K4me1) enhancer signals.

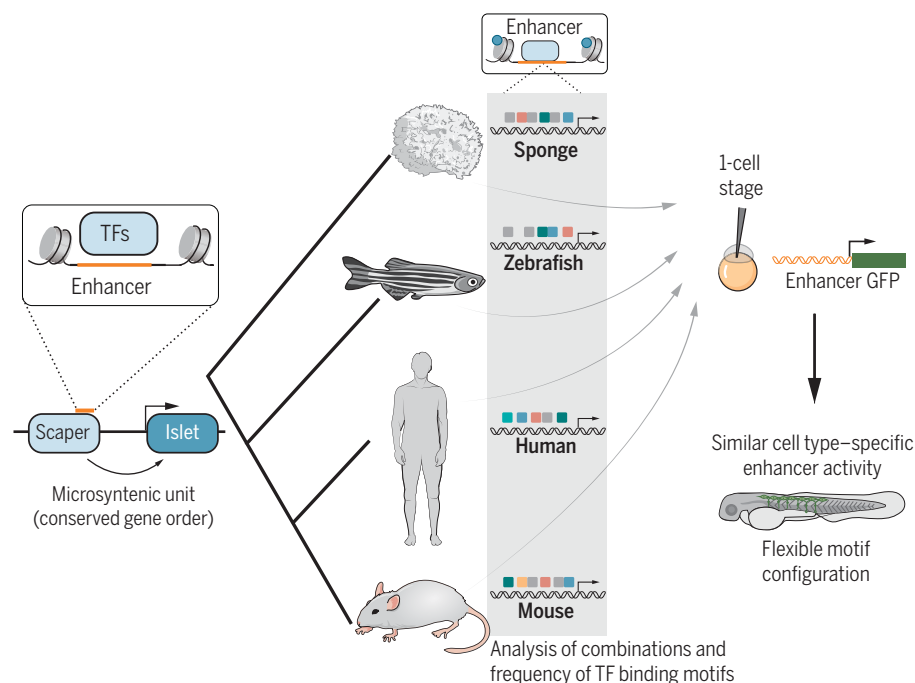
Focusing on an *Islet* enhancer in the *Islet-Scaper* microsyntenic region, we found that the sponge 709-base pair enhancer, independent of its orientation, drives green fluorescent protein (GFP) expression in zebrafish cells in the hind-brain neuroepithelial region, the roof plate around the midline, the pectoral fin, and the otic vesicle; the activity overlaps with endogenous *Isl2a* expression. Systematic removal of sequences from the *Amphimedon Islet* enhancer revealed that both the 5' and 3' regions of this enhancer are required for consistent cell type-specific activity in zebrafish.

We then used the number and frequency of TF binding motifs in the *Amphimedon Islet* enhancer to identify putative enhancers in human, mouse, and fly *Islet-Scaper* regions. The candidate orthologous enhancers from humans and mice drove gene expression patterns similar to those in sponges and endogenous *Islet* enhancers in zebrafish.

We also demonstrated that a number of putative *Amphimedon* enhancers, which are outside conserved microsyntenic regions, can also drive unique expression patterns: Enhancers of sponge housekeeping genes drive broader expression patterns in zebrafish.

CONCLUSION: These results suggest the existence of an ancient and conserved, yet flexible, genomic regulatory syntax that (i) can be interpreted by the available TFs present in cells constituting disparate developmental systems and cell types, and (ii) has been repeatedly co-opted into cell type-specific networks across the animal kingdom.

This common regulatory code maintains a repertoire of conserved TF binding motifs that stabilize and preserve enhancer functionality over evolution. Once established, these enhancers may be maintained as part of conserved gene regulatory network modules over evolution. Although robust, these enhancers can evolve through the expansion and integration of new TF binding motifs and the loss of others. We posit that the expansion of TFs and enhancers may underlie the evolution of complex body plans. ■



***Islet* enhancer activity is conserved across animal evolution.** Enhancers located within conserved microsyntenic units in the sponge *Amphimedon queenslandica* are tested in a zebrafish transgenic reporter system. In zebrafish, the sponge *Islet* enhancer drives a GFP reporter expression pattern similar to that of human, mouse, and zebrafish enhancers identified within the *Islet-Scaper* microsyntenic region. This suggests the conservation of regulatory syntax specified by flexible organizations of motifs.

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

^{*}Corresponding author. Email: e.wong@victorchang.edu.au (E.S.W.); m.francois@centenary.org.au (M.F.); b.degnan@uq.edu.au (B.M.D.)

[†]These authors contributed equally to this work.

[‡]These authors contributed equally to this work.

Cite this article as E. S. Wong et al., *Science* 370, eaax8137 (2020). DOI: 10.1126/science.aax8137

READ THE FULL ARTICLE AT
<https://doi.org/10.1126/science.aax8137>

RESEARCH ARTICLE

ENHANCER GENOMICS

Deep conservation of the enhancer regulatory code in animals

Emily S. Wong^{1,2,3*}, Dawei Zheng^{2†}, Siew Z. Tan^{4†}, Neil L. Bower⁴, Victoria Garside⁵, Gilles Vanwalleghem⁶, Federico Gaiti¹, Ethan Scott⁶, Benjamin M. Hogan^{4,7}, Kazu Kikuchi², Edwina McGlinn⁵, Mathias Francois^{4,8*}†, Bernard M. Degnan^{1*}†

Interactions of transcription factors (TFs) with DNA regulatory sequences, known as enhancers, specify cell identity during animal development. Unlike TFs, the origin and evolution of enhancers has been difficult to trace. We drove zebrafish and mouse developmental transcription using enhancers from an evolutionarily distant marine sponge. Some of these sponge enhancers are located in highly conserved microsyntenic regions, including an *Islet* enhancer in the *Islet-Scaper* region. We found that *Islet* enhancers in humans and mice share a suite of TF binding motifs with sponges, and that they drive gene expression patterns similar to those of sponge and endogenous *Islet* enhancers in zebrafish. Our results suggest the existence of an ancient and conserved, yet flexible, genomic regulatory syntax that has been repeatedly co-opted into cell type-specific gene regulatory networks across the animal kingdom.

Metazoan development and cell type diversity require networks of genome-encoded information. These gene regulatory networks consist of regions of DNA—promoters and enhancers—that contain short sequence motifs [<10 base pairs (bp)] to which transcription factors (TFs) bind. Although promoters and proximal enhancers located adjacent to the start of gene transcription predate animals, distal enhancers marked by H3K4me1 (histone H3 Lys⁴ methylation) or H3K27ac (histone H3 Lys²⁷ acetylation) appear to be a metazoan innovation (1–3) (fig. S1).

Tissue and cell type gene expression profiles are highly conserved across closely related species (e.g., mammals) (4). However, the enhancers controlling gene expression appear to be largely species-specific (5–8), with TF binding sites evolving at a faster rate than the TFs they interact with and the genes they regulate (9). Genomic locations of TF-DNA binding events are highly divergent between mammals (e.g., human and mouse) and are a major source of binding differences between humans (5, 10).

For instance, only ~5% of known TF binding sites used in embryonic stem cells are conserved between humans and mice (11), and only 1% of human tissue-specific enhancers are conserved across mammalian genomes (6). Enhancer similarities at larger evolutionary distances are even more rare. Between Bilateria and its sister lineage Cnidaria, only a single conserved enhancer has been uncovered (12). Thus, despite TF families being highly conserved across the animal kingdom (12, 13) (fig. S1), pan-metazoan enhancers appear to be rare and have yet to be described.

The conservation of syntenic regions in animal genomes has been attributed to constraints imposed by enhancers and other regulatory elements (14–16). These conserved syntenic regions, which are often within topologically associating domains (TADs) in the human genome, comprise cis-regulatory blocks, where enhancers controlling the expression of one gene are located in another gene in the syntenic region (17–19). Almost 600 conserved gene pairs (microsyntenic regions) have been identified across the Metazoa, including vertebrates, tunicates, and arthropods, with 60 to 80 reported in the genome of the marine sponge *Amphimedon queenslandica*, a species representing one of the oldest surviving phyletic lineages of animals (2, 16). Genes within microsyntenic regions are classified into two types: the “bystander” gene, which harbors enhancers that regulate the expression of the “target” gene, which itself is often a developmental regulatory gene (Fig. 1A). The maintenance of this target-bystander microsyntenic unit across animal evolution appears to be attributable to the interleaving regulatory structure (Fig. 1B). Uncoupling or interruption of the bystander gene results in the misregulation of the target gene

(20). Conserved target gene enhancers located in bystander genes remain largely uncharacterized across phyla because of the rapid evolution of these noncoding sequences.

Although enhancer sequences evolve rapidly, they can operate with substantial sequence flexibility, which suggests that regulatory conservation may be more common than has been believed (21). Numerous studies have shown that strict and specific TF binding motif positional conservation is not a requirement for enhancer activity (22–24). Between distant fly species, *eve* stripe 2 enhancer sequences have extensive flexibility in TF binding motif arrangement and spacing, yet yield the same expression pattern (22). Similarly, extreme sequence divergence has been reported in orthologous regulatory regions that produce similar regulatory output in yeast (25), and individual TF binding motifs in yeast and mouse have been shown to tolerate up to 72% of all possible mutations (26).

Characterization of sponge enhancers within conserved microsyntenic regions

Amphimedon enhancers were identified by genome-wide chromatin immunoprecipitation (ChIP) of antibodies for the histone modifications H3K4me1, H3K27ac, H3K4me3 (histone H3 Lys⁴ trimethylation), H3K36me3, and H3K27me3 followed by high-throughput sequencing (ChIP-seq) and alignment of the short reads to the *Amphimedon* genome (2, 13). Of 60 conserved microsyntenic regions identified in *Amphimedon*, 52 contain at least one histone mark specifying enhancer activity in larval or adult stages (e.g., Fig. 1C) (2). According to BLAST search (27), these putative sponge enhancers do not share significant sequence identity with known vertebrate genomes. Of these 52, we selected enhancers located within three conserved microsyntenic gene pairs for further analysis: *Islet-Scaper*, *Tdrd3-Diaph3*, and *Ccne1-Uri* (Fig. 1C and fig. S2). These were chosen according to an unambiguous gene orthology, the strength of the H3K4me1 enhancer signal, and the known regulatory roles of the developmental genes in these microsyntenic regions.

We compared the gene expression profile of microsyntenic gene pairs to determine whether they have distinct expression profiles, using previously published CEL-Seq2 datasets from different cell types and across *Amphimedon* development (28, 29). In all cases, we found that the *Amphimedon* target and bystander genes have uncoupled expression profiles (Fig. 1D and fig. S3, C and D), akin to the uncorrelated expression between target and bystander genes in humans and zebrafish (16, 20).

We then used position-weighted matrices (PWMs) to identify TF binding motifs in these putative *Amphimedon* target enhancers and to

¹School of Biological Sciences, University of Queensland, Brisbane, Australia. ²Victor Chang Cardiac Research Institute, Sydney, Australia. ³School of Biotechnology and Biomolecular Sciences, UNSW Sydney, Sydney, Australia. ⁴Institute for Molecular Biosciences, University of Queensland, Brisbane, Australia. ⁵Australian Regenerative Medicine Institute, Monash University, Melbourne, Australia. ⁶Queensland Brain Institute, University of Queensland, Brisbane, Australia. ⁷Department of Anatomy and Neuroscience and Sir Peter MacCallum Department of Oncology, University of Melbourne, Melbourne, Australia. ⁸Centenary Institute, David Richmond Program for Cardio-Vascular Research: Gene Regulation and Editing, School of Life and Environmental Sciences, University of Sydney, Sydney, Australia.

*Corresponding author. Email: e.wong@victorchang.edu.au (E.S.W.); m.francois@centenary.org.au (M.F.); b.degnan@uq.edu.au (B.M.D.)

†These authors contributed equally to this work.

‡These authors contributed equally to this work.

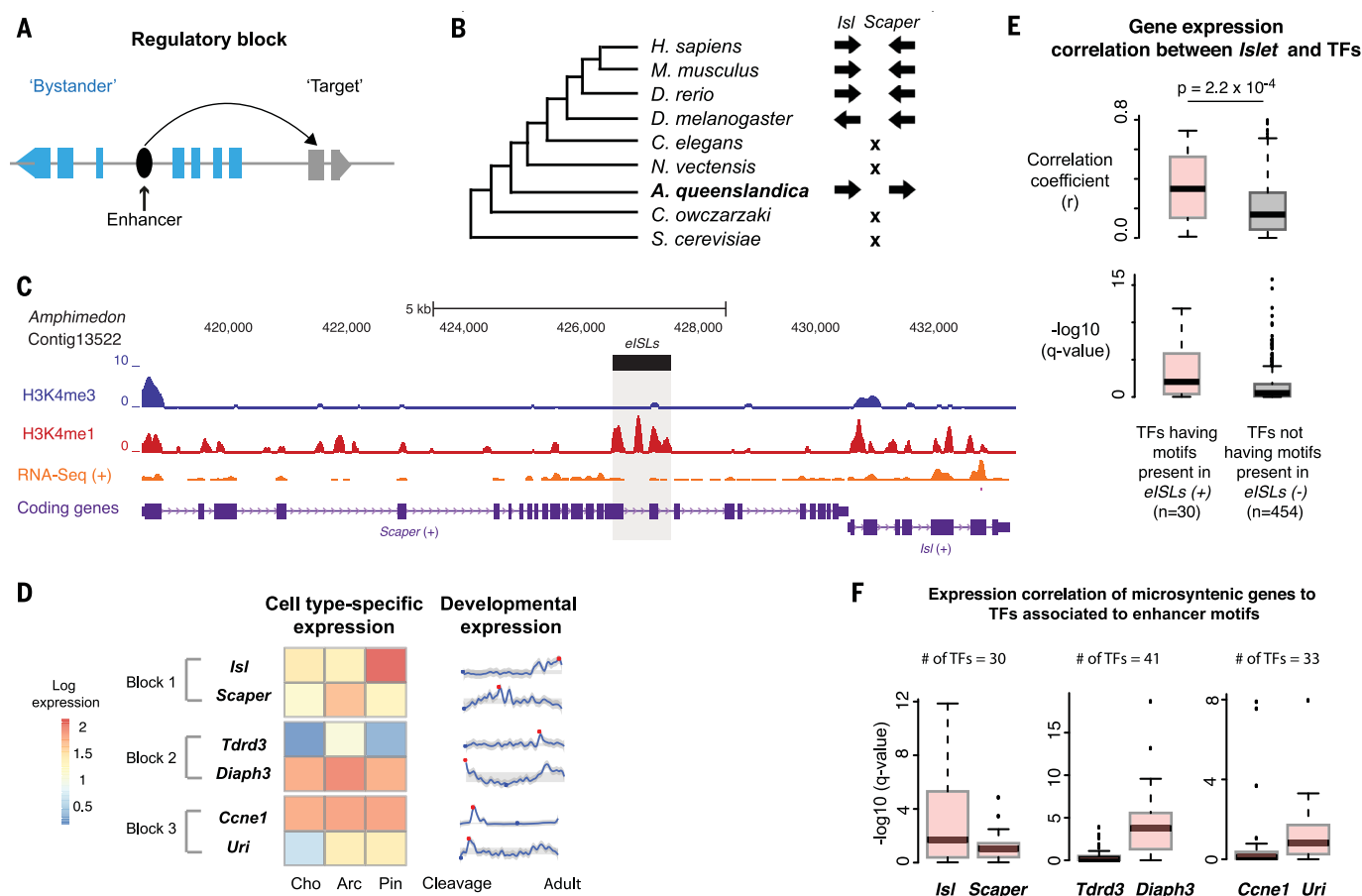


Fig. 1. Microsyntenic regions in *Amphimedon* have regulatory features similar to those of vertebrates. (A) A microsyntenic unit with regulatory enhancers for the target gene, which is often a developmental regulatory gene, residing in the adjacent bystander gene (20). (B) Gene arrangement of the *Islet-Scaper* microsyntenic unit in select animal genomes. Arrows indicate direction of transcription; x, *Islet-Scaper* synteny not present. (C) The location of a putative enhancer in *Amphimedon Scaper* (gray) overlaps with a H3K4me1 peak; adult RNA-seq is shown. H3K27ac is not active in this region (2). (D) Expression levels of genes (\log_{10}) constituting three microsyntenic gene pair blocks in adult *Amphimedon* cell types and during development from early embryo to adult (28, 29) (fig. S3C). Cho, choanocytes; Arc, archaeocytes; Pin, pinacocytes; red and blue dots, maximum and minimum developmental expression levels, respectively;

gray region, interquartile range. (E) Developmental gene expression levels of *Amphimedon Islet* and TF genes whose binding motifs are present in the *eISLs* enhancer ($n = 30$) are more correlated than other TF genes ($n = 454$) (Mann-Whitney U , $P = 2.7 \times 10^{-3}$). (F) Predicted TFs binding to sponge *eISLs*, *eTDRs*, and *eCCNs* enhancers (number at the top of each plot) are better correlated with *Islet*, *Diaph3*, and *Uri* expression than with *Scaper* (Mann-Whitney U , $P = 1.2 \times 10^{-2}$, *Tdrd3* ($P = 2.6 \times 10^{-9}$), and *Ccne1* ($P = 1.8 \times 10^{-3}$) expression, respectively. In (E) and (F), tops and bottoms of boxes denote the 25th and 75th percentile; upper and lower whiskers extend to the largest and smallest value, respectively, no further than 1.5 times the interquartile range. $P < 0.0001$ for sponge *Islet* injected versus uninjected and -RT. Zebrafish *Islet* expression in injected versus uninjected $P = 0.2$, 0.5 , and 0.1 for *Isl1*, *Isl2a*, and *Isl2b*, respectively.

determine whether the expression of the TFs predicted to bind to these motifs covaried with the expression of the genes being regulated by these enhancers. Using this approach, we determined that 117, 160, and 131 motifs are present in *Islet-Scaper* ("eISLs"), *Tdrd3-Diaph3* ("eTDRs"), and *Ccne1-Uri* ("eCCNs") enhancers, respectively (see materials and methods, Fig. 1E, and fig. S4). The expression of TFs identified by their binding motifs in the enhancer of each microsyntenic gene pair was significantly more predictive of the gene expression profile of one gene in the microsyntenic region relative to the other (Mann-Whitney U test, *Islet-Scaper*, $P = 1.2 \times 10^{-2}$; *Ccne1-Uri*, $P = 1.8 \times 10^{-3}$; *Tdrd3-Diaph3*, $P = 2.6 \times 10^{-9}$; Fig. 1F and fig. S4). For example, TFs associated

with the *eISLs* enhancer located within the *Scaper* gene were more strongly correlated with *Islet* expression than with *Scaper*. These results demonstrate that PWMs can meaningfully identify TFs that likely regulate gene expression in *Amphimedon* without prior evidence of the TFs recruited to the putative enhancer sequence.

Sponge enhancers direct tissue-specific expression in vertebrates

To test for functional conservation of *eISLs*, we first generated transgenic zebrafish embryos with the entire *Amphimedon Islet-Scaper* microsyntenic region (15 kb) and tested for sponge gene expression. *Amphimedon Islet* transcripts, but not *Scaper* mRNA, were detected at 48 hours

post-fertilization (hpf) in the transgenic zebrafish larvae (Fig. 2A and fig. S5). The zebrafish *Islet* paralogs *Isl1* (*Isl1*), *Isl2a* (*Isl2a*), and *Isl2b* (*Isl2b*) were also expressed at this stage (Fig. 2A). This result shows that cis-regulatory DNA sequences in the sponge *Islet-Scaper* locus are recognized by zebrafish TFs, resulting in the transcription of the sponge *Islet* gene.

We next assessed the tissue-specific activity of sponge *eISLs*, *eTDRs*, and *eCCNs* enhancer-containing regions in transgenic zebrafish by inserting the enhancer upstream of a silent *gata2a* minimum promoter and a green fluorescent protein (GFP) reporter gene within a zebrafish enhancer detection (ZED) vector (30). This enhancer reporter system is routinely used in zebrafish to detect enhancer activity

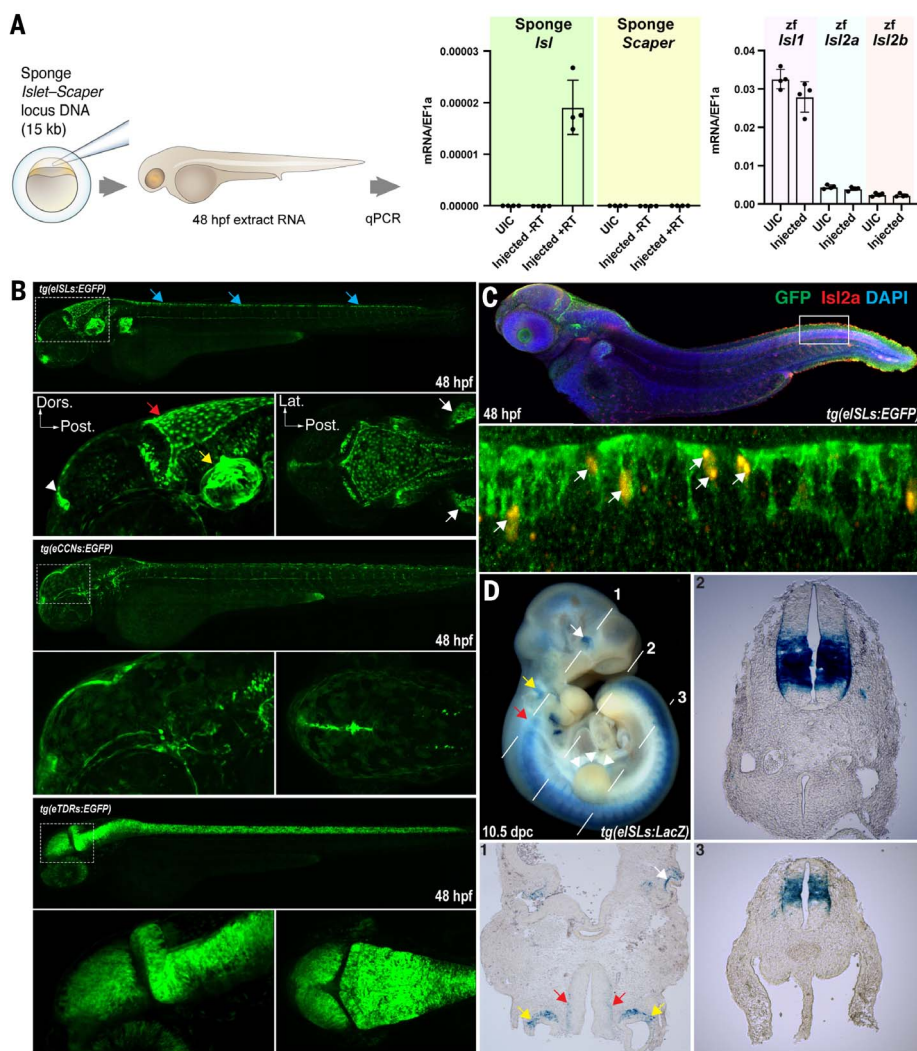


Fig. 2. Sponge enhancers direct cell type-specific expression in vertebrate embryos. (A) Quantitative real-time polymerase chain reaction analysis of sponge gene expression in transient transgenic zebrafish injected with the *Amphimedon Islet-Scaper* microsyntenic locus (15 kb). *Islet* and *Scaper* expression are normalized to *Efla* in uninjected (UIC) zebrafish embryos, injected but not reverse-transcribed (–RT), and injected and reverse-transcribed (+RT). *Amphimedon Islet* expression is present but lower than endogenous zebrafish (*zf*) *Isl1*, *Isl2a* (linked to zebrafish *Scaper*), and *Isl2b* expression. Error bars denote SD. $P < 0.0001$ for sponge *Isl* injected versus uninjected and –RT. Zebrafish *Isl* expression in injected versus uninjected $P = 0.2$, 0.5 , and 0.1 for *Isl1*, *Isl2a*, and *Isl2b*, respectively. (B) Sponge enhancers drive a GFP reporter gene in a zebrafish enhancer trap assay. Stable (F_1) transgenic lines were established for sponge *eISLs*, *eTDRs*, and *eCCNs* enhancers; images show enhancer-driven GFP expression in 48-hpf zebrafish larvae. *eISLs* directs expression in the neuroepithelial layer of the hindbrain (red arrow), the roof plate (blue arrows), the otic vesicle (yellow arrow), the pineal region (arrowhead), and the pectoral fin (white arrows). *eCCNs*-driven GFP expression is similar to that of *eISLs* but not identical, with some additional signal in the periorbital area. *eTDRs* directs pan-neural activity in the spinal cord and in the midbrain. (C) Top: Transgenic fish showing overlapping *eISLs* GFP immunofluorescence (green) and endogenous *Isl2a* in situ hybridization (red) patterns at 48 hpf. Bottom: A subpopulation of neurons from the roof plate expressing both *eISLs*-directed GFP and endogenous *Isl2a* mRNA (arrows). DAPI (4',6-diamidino-2-phenylindole; blue) reveals overall tissue organization. (D) The sponge *eISLs* enhancer drives LacZ reporter expression in whole-mounted 10.5-dpc mouse embryos (blue β -Gal staining) in transient transgenic mice. Expression in the neural tube (red arrows), apical-ectodermal ridge of the limb bud (arrowheads), otic vesicle (yellow arrows), and the inner layer of the optic cup (white arrows) is detected in three of five transgenics. Images of transverse sections (anatomical location is indicated by white dashed lines 1, 2, and 3) show specific activity of *eISL* enhancer in a subset of neurons in the neural tube along an anteroposterior axis.

in vivo (31, 32), with reporter expression reflecting that of the endogenous gene (33).

All three *Amphimedon* enhancers drove specific reporter expression during early zebrafish embryo development (Fig. 2B and fig. S8). The 709-bp *eISLs* enhancer-containing region is active in the hindbrain neuroepithelial region, a subset of sensory neurons in the roof plate around the midline as well as in the pectoral fin and otic vesicle. Expression patterns are consistent regardless of enhancer orientation, supporting the notion that enhancers are typically flexible in their orientation to the gene transcribed (34) (fig. S9). Similarly, the 931-bp *eCCNs* and the 1510-bp *eTDRs* enhancer regions are active in neuronal lineages (Fig. 2B and fig. S8). *eTDRs* shows a pan-neural activity, labeling most neurons from the spinal cord and mid-brain region, and *eCCNs* is active in a small periorbital neuron population. *eCCNs* also shows a low degree of penetrance in the tail in additional founder lines (Fig. 2B and fig. S6). Independent stable transgenic lines were generated for each construct (Fig. 2B and fig. S6). GFP signal was not detected in animals injected with empty vectors containing the minimal promoter and GFP sequence (fig. S7). To confirm that the *Amphimedon* sequence was driving reporter expression, we excised the inserted *eISLs* sequence from stable *eISLs* fish lines. Independent reporter fish lines were injected at one-cell stage with *iCre* mRNA to remove the *eISLs* region. A marked reduction of GFP expression occurred in the *iCre*-injected larvae relative to the *iCre*-negative animal from the same founder line (fig. S10).

We performed RNA sequencing (RNA-seq) on GFP-positive and -negative cell fractions from stable fish reporter lines and mapped functional terms associated with significantly up-regulated genes in reporter-positive cells (fig. S11). This analysis confirmed that each sponge enhancer labels specific cell populations consistent with observed anatomical location. For example, genes linked to otolith, notochord, melanocyte, and fin development were enriched in *eISLs*-positive cells, supporting reporter gene localization in the ear, spinal cord, skin, and pectoral fin, respectively ($P < 0.05$, more than 5 times greater) (see also fig. S12).

To determine whether the *Amphimedon eISLs* enhancer was also active in the same cells and at the same time as the endogenous zebrafish *Isl2a* transcript, we combined *Isl2a* fluorescence in situ hybridization with GFP immunofluorescence in 48-hpf *eISLs* reporter zebrafish embryos. We found that endogenous *Isl2a* expression is consistent with previously published expression patterns (35) and overlaps with *eISLs*-driven GFP expression. The exogenous sponge enhancer also is active in additional cell types, including in distinct sensory neurons in the roof plate (Fig. 2C, white arrows). The broader GFP expression pattern relative

to the endogenous *Isl2a* expression profile suggests that the sponge *eISLs* enhancer either possesses additional regulatory sequences for TFs not involved in normal zebrafish *Isl2a* regulation or lacks regulatory elements that restrict endogenous activity to the *Isl2a*-positive cells (Fig. 2C).

To determine whether the same sponge enhancers have similar cell type-specific activity in mammals, we characterized the expression of LacZ reporter constructs in transient transgenic mouse embryos at 10.5 days post coitum (dpc) (Fig. 2D and fig. S13). The *Amphimedon eISLs*, *eTDRs*, and *eCCNs* enhancers direct β -galactosidase activity in a tissue-specific manner in mice. These activity patterns are similar to the expression patterns observed in fish reporter assays (Fig. 2B). For instance, β -galactosidase activity for *eISLs* is present in the mouse neural tube, lens, otic vesicle, and limb bud, which suggests that the sponge enhancers are interpreted by conserved tissue-specific gene regulatory networks in vertebrates.

Given that the tested sponge microsyntenic-derived enhancers all appear to direct gene expression in vertebrates, we sought to determine whether enhancers in nonconserved genomic regions in the sponge also function in vertebrates. We selected two *Amphimedon* enhancers on the basis of histone marks located within protein-coding genes but not in a metazoan microsyntenic region (termed “*eIs*” and “*e2s*”) (16) (fig. S3, C and D, and fig. S14). *eIs* is located within a gene similar to *DLG5*, a scaffold protein in humans (Aqu2.1.24542, blastp $e = 1.6 \times 10^{-6}$); *e2s* is located in a sponge-specific gene without obvious similarity to other known genes (Aqu2.1.25069, blastp $e < 10$, nr database). Relative to enhancers derived from conserved microsyntenic units, we found that these intragenic enhancers, in particular *eIs*, drive broad variable reporter expression in stable fish lines and transient transgenic reporter mouse embryos (figs. S15 and S16). The gene harboring *eIs*, Aqu2.1.24542, is highly expressed in *Amphimedon* adult cells and across development, similar to bystander genes (fig. S4, B and C). This suggests that *eIs* is not developmentally regulated and supports the proposition that metazoan enhancers have either developmental or housekeeping functions (36).

In summary, despite a lack of primary sequence similarity, *Amphimedon* enhancers within microsyntenic regions are active in teleost and mammalian development and direct cell type-specific activity. Interestingly, sponge enhancers located outside of conserved microsyntenic regions, and which likely regulate nondevelopmental genes, also are able to respond to TF activation in vertebrates, even though their last common ancestor lived hundreds of millions of years ago. This suggests that the combinations of TF binding sites in sponge enhancers can be broadly interpreted across the animal kingdom.

Identification of putative bilaterian orthologs of a sponge enhancer

To gain insight into the sequence components driving *eISLs* activity in zebrafish, we performed a series of enhancer truncation experiments based on TF binding motif location on *eISLs*. These motifs fall into four broad clusters (Fig. 3A). We generated multiple stable founder lines deleting four regions between 149 and 180 bp in length (termed D1 to D4). With the exception of the D3 deletion, all enhancer truncation lines are markedly different from the lines with the full *eISLs* enhancer at 48 hpf (Fig. 3A and fig. S17). D1 removal reveals highly variable expression patterns both within and between founder lines, suggestive of key sequence elements that are required to maintain enhancer specificity. The D2 deleted lines show a marked decrease in the strength of reporter expression, consistent with the removal of a high number of TF binding motifs present in this region (fig. S18 and table S12). Deletion of D4 results in a decrease of neuronal expression in the spinal region in three of four lines. These results suggest that multiple sequences across *eISLs* are responsible for cell type-specific activity.

Sponges do not possess eumetazoan cell types (e.g., neurons, muscles), with the possible exception of epithelial cells. Given the ability of sponge enhancers to drive cell type-specific expression in zebrafish despite ~700 million years of independent evolution and a lack of common cell types, we sought to determine whether the *Amphimedon* enhancers possess a conserved cis-regulatory signature that could be used to identify similar enhancers in bilaterians despite the absence of detectable primary sequence similarity.

We devised a method to quantify the set of TF binding motifs in an enhancer to allow alignment with divergent enhancers. Specifically, we calculated a score using the PWM alignment score and the frequency of the motif that was independent of the relative spacing and order of the motifs with respect to each other (Fig. 3B). We then scanned bilaterian microsyntenic regions (target) with sponge enhancer sequences (query) and measured sequence similarity using a sliding-window approach. We assumed that for enhancers to be maintained across large evolutionary distances, their regulatory syntax must allow for extensive flexibility in motif ordering and spacing.

Because the candidate zebrafish enhancer of *Isl2a*, which is localized to a *Scaper* intron and drives an expression pattern similar to the endogenous *Isl2a* transcript, has already been characterized (16) (fig. S19), we applied this approach to mammals and *Drosophila*. Using the query output from the *Amphimedon eISLs* enhancer, we first tested this approach on a 557-kb *Islet-Scaper* microsyntenic sequence in humans. A 900-bp human region (chr15: 76358600 to 76359500) best matches the sponge

query and aligns to the mouse genome near the 3' end of the *Scaper* gene in a region of conserved exon synteny to the *Isl2a* candidate enhancer in zebrafish. This region is a potential enhancer in mice, marked by H3K4me1 (10.5 dpc) in the face and midbrain (ENCODE: Encyclopedia of DNA Elements), where only a small fraction of the total mouse *Islet-Scaper* locus shows the H3K4me1 mark (face and midbrain H3K4me1 marks span only 0.02% and 0.01% of the total *Islet-Scaper* locus, respectively). Predicted human and mouse orthologs of the other two sponge enhancers, *eCCNs* and *eTDRs*, have been identified using this approach and can be found in the supplementary materials.

Although comparisons of putative sponge, fish, human, and mouse *eISL* enhancers revealed high variability in the composition, alignment, and frequency of known TF binding motifs (Fig. 3C and figs. S20 and S21), there are 39 binding motifs common to all species (Fig. 3C, accounting for TF family expansions). Among these species, there is high flexibility in the type and spacing of motifs; 29 are single TF binding sites and 10 are composite motifs. TF binding sites span 13 classes of TFs, including binding motifs for basic leucine zipper factors (bZIP), motifs for AP-1 and C/EBP, and NK homeodomain proteins; these TF families are present in *Amphimedon* (37, 38).

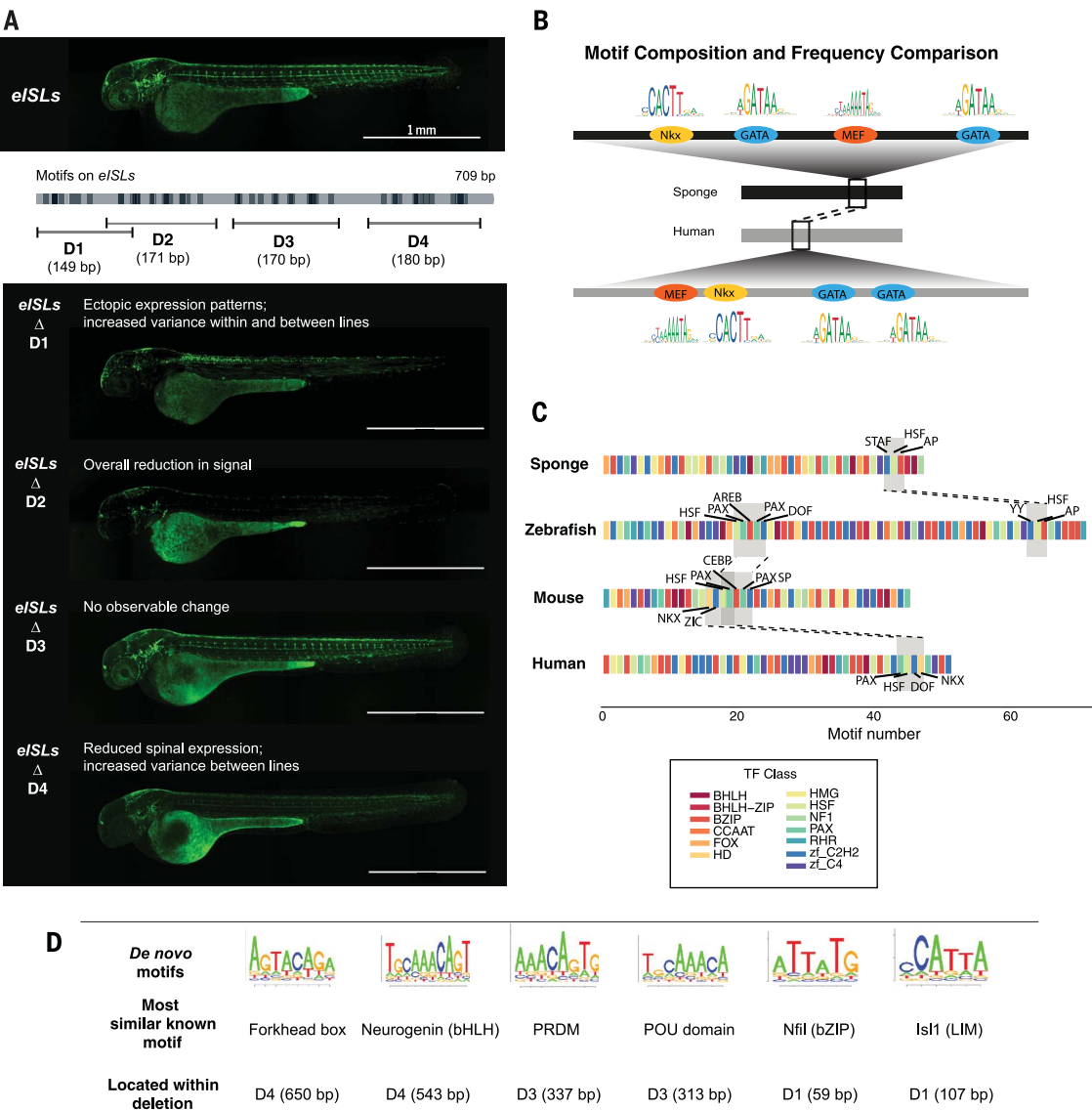
Using a de novo pattern identification method to identify motifs shared by the *eISL* enhancer sequences from each species (39), we detected seven motifs and matched these to six known vertebrate motifs (Fig. 3D). Notably, a motif recognized by human *Neurogenin-2*, a basic helix-loop-helix (bHLH) TF that is active in the neural tube and involved in neuron differentiation, is present. This TF binds to enhancer-box elements in the promoter of numerous genes involved in neurogenesis, and it forms a complex involving *Islet1*—the LIM-homeodomain transcription complex—that specifically drives motor neuron fate (40). When the sequence harboring this motif in the *Amphimedon eISLs* is deleted (D4), expression in the spinal region is markedly reduced (Fig. 3A). Interestingly, the *Amphimedon* ortholog of *Neurogenin*, *AmqbHLH1*, is expressed in putative larval sensory cells and has strong proneuronal activity when expressed in *Drosophila* and *Xenopus* embryos (41).

The relative position of TF binding sites along an enhancer can be critical, as this can change the secondary structure of the DNA and affect the binding affinity of cofactor proteins (42–44). Given this, we sought to determine whether motif order between sponge and vertebrate enhancers is conserved by pairwise sequence alignment of sequences based on motifs. We found evidence for motif order conservation by global-local independent alignment between the sponge and human enhancers

Fig. 3. Identification of vertebrate enhancers orthologous to *Amphimedon* *eISLs* based on TF motif composition. (A) Top:

GFP expression of the full *eISLs* enhancer. Middle: GFP reporter activity in transgenic zebrafish with *eISLs* enhancer deletions. The gray bar represents the full *eISLs* sequence; high-scoring TF binding motifs are shown as dark bars, and deleted regions D1 to D4 are indicated by brackets. Bottom: Representative GFP expression profiles of F₁ transgenic zebrafish for each of the deleted *Amphimedon* *eISLs*. (B) A simple representation of the motif composition and frequency “bag of motifs” method for computing similarity between divergent enhancer sequences on the basis of type and frequency of TF binding motifs from PWMs (see supplementary materials). (C) Locations of TF motifs are displayed according to their TF class for each *eISL* enhancer in sponge, zebrafish, mouse, and human (53). Gray regions connected by dashed lines are locally aligned at *P* < 0.1. Only shared motifs among sequences are depicted.

(D) Representative sequence logos of de novo identified motifs shared across sponge, zebrafish, human, and mouse *eISL* enhancers. The best-matching TF binding motifs that matched to the JASPAR CORE vertebrate database are shown, together with the locations of these in the *Amphimedon* *eISLs* and the corresponding deleted region.



and between the sponge and mouse enhancers (*P* = 0.02 and *P* = 0.04, respectively; *P* value from empirical null distributions by scrambling the order of motifs). Aligned regions were short, comprising two or three motifs, and the best-aligned regions between sponge and human did not match the best-aligned regions in the sponge-mouse alignment (fig. S21). Using a more relaxed model based on the binding sites of 13 TF classes instead of specific TFs, we did not detect evidence of conserved motif order across all pairwise comparisons. Further, *Scaper* exon-intron structure and *eISLs* enhancer location within *Scaper* were not conserved between sponges and vertebrates, despite conservation of the *Scaper* coding sequence. To determine whether *Islet* enhance conservation extends beyond vertebrates and

sponges, we identified the best-aligned region in *Drosophila* using *eISLs* and the motif-based approach described above, targeting the fly *Isl-Scaper* microsyntenic region. The identified region—an 800-bp sequence within the *Drosophila Scaper* ortholog, *ssp3* [chr2L: 18920300 to 18921100 (dm6), 35 bins; fig. S22]—shows chromatin accessibility in cells of the central nervous system, where the fly *Islet* ortholog (*tup*) is expressed (45), and is located in the genome at the 3' end of *ssp3*, similar to vertebrate *eISLs* in *Scaper*. **Mammalian and sponge *Islet* enhancers drive similar expression in zebrafish** To assess the potential homology of functional activity between mammal and sponge enhancers, we made stable fish lines for the

predicted human and mouse *ISL* enhancers. Cross-species comparisons revealed that all enhancers drive reporter gene expression in the sensory neurons of the roof plate as well as in the hindbrain neuroepithelial layer, the pineal region, and the ear capsule (Fig. 4, A to C, and fig. S23). The sponge *eISL* shows robust activity in neurons of the pectoral fin, whereas mouse and human *eISL* enhancers appear weak in this particular tissue. Of note, both the mouse and human *eISL* enhancers display specific activity in the eye tissue; mouse *eISL* activity is detected in a polarized population of neurons in the anterior part of the retina, whereas human *eISL* shows activity in the central part of the lens and in the proctodeum. Only the human *eISL* enhancer directs clear non-neural activity, with GFP

expression present in a subset of endothelial cells from the intersomitic vasculature. When sponge and mouse enhancer reporter signals are overlaid in a common reference template, a subset of neurons in which both regulatory regions are active is revealed (Fig. 4C, white signal).

Discussion

Even though sponges and vertebrates diverged around 700 million years ago, regulatory enhancers from *Amphimedon* were able to drive developmental cell type-specific expression in zebrafish and mice. All sponge enhancers tested in this study lack discernable sequence similarity to vertebrate genomes. Although we primarily focused on *Amphimedon* enhancers located in deeply conserved microsyntenic gene regions, intragenic enhancers within nonsyntenic and sponge-specific genes also directed distinct expression patterns in

vertebrates. These results suggest that a common regulatory grammar exists in metazoan enhancer sequences that can be interpreted by the available TFs present in cells constituting disparate developmental systems and body plans.

We suggest that the number, location, and type of TF binding sites can change markedly within a conserved enhancer over evolution without the loss of regulatory functionality and even cell type specificity. The *ISL* enhancers in sponges and vertebrates share a set of TF binding motifs that interact with conserved TFs (e.g., Neurogenin, NK homeodomain, AP-1) (table S11), yet differ in their order and frequency. The presence of common TF binding motifs in enhancers in divergent animals is consistent with some of these motifs being critical for the maintenance of enhancer functionality. The results of systematic deletions of the *Amphimedon eISLs*, coupled to de novo

motif discovery, suggest that the endogenous zebrafish TFs uniquely interact with different parts of the sponge enhancer. This finding, together with the absence of a conserved cis-regulatory motif order, supports a regulatory code where a cooperative binding environment is maintained in animal enhancers despite extensive sequence changes, resulting in a highly evolvable and flexible system that maintains, at its core, the ability of conserved TFs to bind conserved motifs. We propose that a core set of TF-DNA interactions stabilizes and maintains enhancer functionality over evolution. Its flexible organization promotes robustness and may provide a foundation for further lineage-specific elaboration via the integration of additional TF binding motifs and the dissociation of others.

Regulatory co-option is an essential part of organismal evolution (27). Sponge and vertebrate development and body plans have little

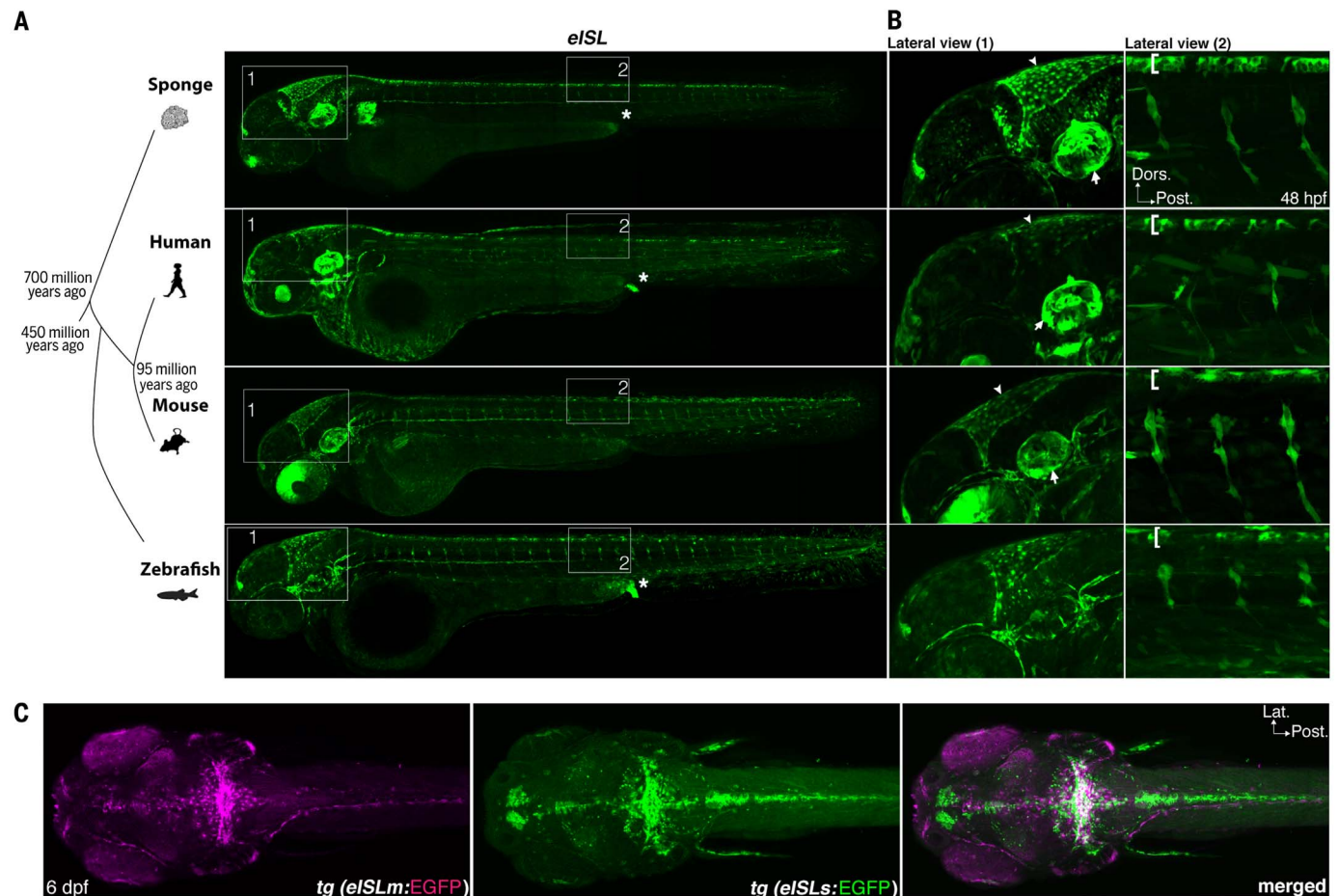


Fig. 4. Mammalian orthologs of the *Amphimedon eISL* enhancer direct similar developmental expression in zebrafish. (A) Comparison of GFP expression pattern at 48 hpf for four stable transgenic zebrafish lines harboring candidate *eISL* enhancers from sponge, human, mouse, and fish. Asterisk indicates proctodeum. (B) Close-up of lateral views from insets numbered 1 and 2 in (A). Lateral view 1 shows the pineal region, the neuroepithelium of the hindbrain (white arrowheads), and the otic vesicle (white arrows). The retina and the lens show expression in

mouse *eISL* and human *eISL*, respectively. In lateral view 2, brackets indicate the floor plate; the intersomitic vessel is evident in human *eISL*. (C) Neuroanatomical visualization of the mouse (*eISLm*) and sponge (*eISLs*) enhancer activities in 6-dpf zebrafish using a common template built from confocal images. Subsets of cells in the caudal part of the cerebellum appear active in mouse *eISL* (purple) and sponge *eISL* (green). White signal in the rightmost ("merged") image consists of overlapping signals from reporters driven by mouse and sponge *eISL*.

in common. Sponges lack muscles, nerves, and a gut. Their cell lineages do not appear to progressively restrict to a germ layer, and they lack a mesoderm; it is questionable as to whether they even gastrulate (46). Thus, the expression patterns generated by the sponge enhancers in fish and mammals, which are largely restricted to neurectoderm lineages, suggest that gene regulatory circuits can be co-opted to perform lineage-specific developmental and cell type gene network functions (47). Given the apparent expansion of the number of enhancers in eumetazoans after diverging from sponges (1, 48–50), which appears to parallel increases in metazoan TF family memberships (51), the duplication and divergence of ancestral enhancers comprising TF binding sites would have yielded an expanded and more complex repertoire of enhancers in vertebrates that, in turn, could be continually co-opted into new functions. We posit that the differential expansion of TFs and enhancers in eumetazoans—and their subsequent co-option into new regulatory roles—underlies the evolution of their complex body plans relative to the simple and restricted body plan of sponges.

Materials and methods summary

Candidate enhancers located within and outside microsyntenic regions in the sponge *Amphimedon queenslandica* were identified on the basis of histone marks (2), and the activity of these sequences was tested in zebrafish and mouse transgenic reporter systems. Tissue-specific activity of sponge enhancers was assessed by imaging stable independent transgenic zebrafish lines using an enhancer reporter vector (30) at different embryonic time points and transgenic mouse embryos with LacZ reporter constructs. Deleted versions of the sponge *Isl* enhancer were tested in stable transgenic zebrafish lines to identify the key sequences driving reporter activity. TF binding motifs within sponge enhancers were identified using PWMs (52). Starting from the analysis of the type and frequency of TF binding motifs in sponge enhancers, we identified candidate orthologous enhancers at the *Isl-Scaper* microsyntenic regions in humans, mice, and *Drosophila*. Human, mouse, and zebrafish *Isl* enhancers were assessed in the zebrafish transgenic reporter system and were compared to the *Amphimedon Isl*-driven reporter expression. Code is available at Zenodo under DOI: 10.5281/zenodo.4001143.

REFERENCES AND NOTES

- M. Schwaiger et al., Evolutionary conservation of the eumetazoan gene regulatory landscape. *Genome Res.* **24**, 639–650 (2014). doi: [10.1101/gr.162529.113](https://doi.org/10.1101/gr.162529.113); pmid: [24642862](https://pubmed.ncbi.nlm.nih.gov/24642862/)
- F. Gaiti et al., Landscape of histone modifications in a sponge reveals the origin of animal cis-regulatory complexity. *eLife* **6**, e22194 (2017). doi: [10.7554/eLife.22194](https://doi.org/10.7554/eLife.22194); pmid: [28395144](https://pubmed.ncbi.nlm.nih.gov/28395144/)
- A. Sebá-Pedros et al., The dynamic regulatory genome of *Capsspora* and the origin of animal multicellularity. *Cell* **165**, 1224–1237 (2016). doi: [10.1016/j.cell.2016.03.034](https://doi.org/10.1016/j.cell.2016.03.034); pmid: [27114036](https://pubmed.ncbi.nlm.nih.gov/27114036/)
- D. Brawand et al., The evolution of gene expression levels in mammalian organs. *Nature* **478**, 343–348 (2011). doi: [10.1038/nature10532](https://doi.org/10.1038/nature10532); pmid: [22012392](https://pubmed.ncbi.nlm.nih.gov/22012392/)
- D. Schmidt et al., Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science* **328**, 1036–1040 (2010). doi: [10.1126/science.1186176](https://doi.org/10.1126/science.1186176); pmid: [20378774](https://pubmed.ncbi.nlm.nih.gov/20378774/)
- D. Villar et al., Enhancer evolution across 20 mammalian species. *Cell* **160**, 554–566 (2015). doi: [10.1016/j.cell.2015.01.006](https://doi.org/10.1016/j.cell.2015.01.006); pmid: [25635462](https://pubmed.ncbi.nlm.nih.gov/25635462/)
- M. J. Blow et al., ChIP-Seq identification of weakly conserved heart enhancers. *Nat. Genet.* **42**, 806–810 (2010). doi: [10.1038/ng.650](https://doi.org/10.1038/ng.650); pmid: [20729851](https://pubmed.ncbi.nlm.nih.gov/20729851/)
- C. D. Arnold et al., Quantitative genome-wide enhancer activity maps for five *Drosophila* species show functional enhancer conservation and turnover during cis-regulatory evolution. *Nat. Genet.* **46**, 685–692 (2014). doi: [10.1038/ng.3009](https://doi.org/10.1038/ng.3009); pmid: [24908250](https://pubmed.ncbi.nlm.nih.gov/24908250/)
- E. S. Wong et al., Decoupling of evolutionary changes in transcription factor binding and gene expression in mammals. *Genome Res.* **25**, 167–178 (2015). doi: [10.1101/gr.177840.114](https://doi.org/10.1101/gr.177840.114); pmid: [25394363](https://pubmed.ncbi.nlm.nih.gov/25394363/)
- M. Kasowski et al., Variation in transcription factor binding among humans. *Science* **328**, 232–235 (2010). doi: [10.1126/science.1183621](https://doi.org/10.1126/science.1183621); pmid: [20299548](https://pubmed.ncbi.nlm.nih.gov/20299548/)
- G. Kunarso et al., Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nat. Genet.* **42**, 631–634 (2010). doi: [10.1038/ng.600](https://doi.org/10.1038/ng.600); pmid: [20526341](https://pubmed.ncbi.nlm.nih.gov/20526341/)
- J. L. Royo et al., Transphylic conservation of developmental regulatory state in animal evolution. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 14186–14191 (2011). doi: [10.1073/pnas.1109037108](https://doi.org/10.1073/pnas.1109037108); pmid: [21844364](https://pubmed.ncbi.nlm.nih.gov/21844364/)
- M. Srivastava et al., The *Amphimedon queenslandica* genome and the evolution of animal complexity. *Nature* **466**, 720–726 (2010). doi: [10.1038/nature09201](https://doi.org/10.1038/nature09201); pmid: [20686567](https://pubmed.ncbi.nlm.nih.gov/20686567/)
- H. Kikuta et al., Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates. *Genome Res.* **17**, 545–555 (2007). doi: [10.1101/gr.6086307](https://doi.org/10.1101/gr.6086307); pmid: [17387144](https://pubmed.ncbi.nlm.nih.gov/17387144/)
- P. G. Engström, S. J. Ho Sui, O. Drivenes, T. S. Becker, B. Lenhard, Genomic regulatory blocks underlie extensive microsynteny conservation in insects. *Genome Res.* **17**, 1898–1908 (2007). doi: [10.1101/gr.6669607](https://doi.org/10.1101/gr.6669607); pmid: [17989259](https://pubmed.ncbi.nlm.nih.gov/17989259/)
- M. Irimia et al., Extensive conservation of ancient microsynteny across metazoans due to cis-regulatory constraints. *Genome Res.* **22**, 2356–2367 (2012). doi: [10.1101/gr.139725.112](https://doi.org/10.1101/gr.139725.112); pmid: [22722344](https://pubmed.ncbi.nlm.nih.gov/22722344/)
- A. Woolfe et al., Highly conserved non-coding sequences are associated with vertebrate development. *PLOS Biol.* **3**, e7 (2005). doi: [10.1371/journal.pbio.0030007](https://doi.org/10.1371/journal.pbio.0030007); pmid: [15630479](https://pubmed.ncbi.nlm.nih.gov/15630479/)
- A. Sandelin et al., Arrays of ultraconserved non-coding regions span the loci of key developmental genes in vertebrate genomes. *BMC Genomics* **5**, 99 (2004). doi: [10.1186/1471-2164-5-99](https://doi.org/10.1186/1471-2164-5-99); pmid: [15613238](https://pubmed.ncbi.nlm.nih.gov/15613238/)
- N. Harmston et al., Topologically associating domains are ancient features that coincide with Metazoan clusters of extreme noncoding conservation. *Nat. Commun.* **8**, 441 (2017). doi: [10.1038/s41467-017-00524-5](https://doi.org/10.1038/s41467-017-00524-5); pmid: [28874668](https://pubmed.ncbi.nlm.nih.gov/28874668/)
- A. Akalin et al., Transcriptional features of genomic regulatory blocks. *Genome Biol.* **10**, R38 (2009). doi: [10.1186/gb-2009-10-4-r38](https://doi.org/10.1186/gb-2009-10-4-r38); pmid: [19374772](https://pubmed.ncbi.nlm.nih.gov/19374772/)
- Y. Yao et al., Cis-regulatory architecture of a brain signaling center predates the origin of chordates. *Nat. Genet.* **48**, 575–580 (2016). doi: [10.1038/ng.3542](https://doi.org/10.1038/ng.3542); pmid: [27064252](https://pubmed.ncbi.nlm.nih.gov/27064252/)
- E. E. Hare, B. K. Peterson, V. N. Iyer, R. Meier, M. B. Eisen, Seisid even-skipped enhancers are functionally conserved in *Drosophila* despite lack of sequence conservation. *PLOS Genet.* **4**, e1000106 (2008). doi: [10.1371/journal.pgen.1000106](https://doi.org/10.1371/journal.pgen.1000106); pmid: [18584029](https://pubmed.ncbi.nlm.nih.gov/18584029/)
- M. Z. Ludwig, N. H. Patel, M. Kreitman, Functional analysis of eve stripe 2 enhancer evolution in *Drosophila*: Rules governing conservation and change. *Development* **125**, 949–958 (1998). pmid: [9449677](https://pubmed.ncbi.nlm.nih.gov/9449677/)
- G. Junion et al., A transcription factor collective defines cardiac cell fate and reflects lineage history. *Cell* **148**, 473–486 (2012). doi: [10.1016/j.cell.2012.01.030](https://doi.org/10.1016/j.cell.2012.01.030); pmid: [22304916](https://pubmed.ncbi.nlm.nih.gov/22304916/)
- D. Zeevi et al., Molecular dissection of the genetic mechanisms that underlie expression conservation in orthologous yeast ribosomal promoters. *Genome Res.* **24**, 1991–1999 (2014). doi: [10.1101/gr.179259.114](https://doi.org/10.1101/gr.179259.114); pmid: [25294245](https://pubmed.ncbi.nlm.nih.gov/25294245/)
- J. L. Payne, A. Wagner, The robustness and evolvability of transcription factor binding sites. *Science* **343**, 875–877 (2014). doi: [10.1126/science.1249046](https://doi.org/10.1126/science.1249046); pmid: [24558158](https://pubmed.ncbi.nlm.nih.gov/24558158/)
- S. F. Altschul et al., Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997). doi: [10.1093/nar/25.17.3389](https://doi.org/10.1093/nar/25.17.3389); pmid: [9254694](https://pubmed.ncbi.nlm.nih.gov/9254694/)
- S. Sogabe et al., Pluripotency and the origin of animal multicellularity. *Nature* **570**, 519–522 (2019). doi: [10.1038/s41586-019-1290-4](https://doi.org/10.1038/s41586-019-1290-4); pmid: [31189954](https://pubmed.ncbi.nlm.nih.gov/31189954/)
- M. Levin et al., The mid-developmental transition and the evolution of animal body plans. *Nature* **531**, 637–641 (2016). doi: [10.1038/nature16994](https://doi.org/10.1038/nature16994); pmid: [26886793](https://pubmed.ncbi.nlm.nih.gov/26886793/)
- J. Bessa et al., Zebrafish enhancer detection (ZED) vector: A new tool to facilitate transgenesis and the functional analysis of cis-regulatory regions in zebrafish. *Dev. Dyn.* **238**, 2409–2417 (2009). doi: [10.1002/dvdy.22051](https://doi.org/10.1002/dvdy.22051); pmid: [19653328](https://pubmed.ncbi.nlm.nih.gov/19653328/)
- I. K.-N. Chiang et al., SoxF factors induce Notch1 expression via direct transcriptional regulation during early arterial development. *Development* **144**, 2629–2639 (2017). doi: [10.1242/dev.146241](https://doi.org/10.1242/dev.146241); pmid: [28619820](https://pubmed.ncbi.nlm.nih.gov/28619820/)
- F. Marletaz et al., Amphioxus functional genomes and the origins of vertebrate gene regulation. *Nature* **564**, 64–70 (2018). doi: [10.1038/s41586-018-0734-6](https://doi.org/10.1038/s41586-018-0734-6); pmid: [30464347](https://pubmed.ncbi.nlm.nih.gov/30464347/)
- S. Ellingsen et al., Large-scale enhancer detection in the zebrafish genome. *Development* **132**, 3799–3811 (2005). doi: [10.1242/dev.01951](https://doi.org/10.1242/dev.01951); pmid: [16049110](https://pubmed.ncbi.nlm.nih.gov/16049110/)
- A. Hozumi et al., Enhancer activity sensitive to the orientation of the gene it regulates in the chordate genome. *Dev. Biol.* **375**, 79–91 (2013). doi: [10.1016/j.ydbio.2012.12.012](https://doi.org/10.1016/j.ydbio.2012.12.012); pmid: [23274690](https://pubmed.ncbi.nlm.nih.gov/23274690/)
- R. L. Moreno, K. Williams, K. L. Jones, A. B. Ribera, Investigation of *Islet2a* function in zebrafish embryos: Mutants and morphants differ in morphologic phenotypes and gene expression. *PLOS ONE* **13**, e0199233 (2018). doi: [10.1371/journal.pone.0199233](https://doi.org/10.1371/journal.pone.0199233); pmid: [29927984](https://pubmed.ncbi.nlm.nih.gov/29927984/)
- M. A. Zabidi et al., Enhancer-core-promoter specificity separates developmental and housekeeping gene regulation. *Nature* **518**, 556–559 (2015). doi: [10.1038/nature13994](https://doi.org/10.1038/nature13994); pmid: [25517091](https://pubmed.ncbi.nlm.nih.gov/25517091/)
- K. Jindrich, B. M. Degnan, The diversification of the basic leucine zipper family in eukaryotes correlates with the evolution of multicellularity. *BMC Evol. Biol.* **16**, 28 (2016). doi: [10.1186/s12862-016-0598-z](https://doi.org/10.1186/s12862-016-0598-z); pmid: [26831906](https://pubmed.ncbi.nlm.nih.gov/26831906/)
- C. Larroux et al., The NK homeobox gene cluster predates the origin of Hox genes. *Curr. Biol.* **17**, 706–710 (2007). doi: [10.1016/j.cub.2007.03.008](https://doi.org/10.1016/j.cub.2007.03.008); pmid: [17379523](https://pubmed.ncbi.nlm.nih.gov/17379523/)
- G. Pavesi, P. Mereghetti, G. Mauri, G. Pesole, Weeder Web: Discovery of transcription factor binding sites in a set of sequences from co-regulated genes. *Nucleic Acids Res.* **32**, W199–W203 (2004). doi: [10.1093/nar/gkh465](https://doi.org/10.1093/nar/gkh465); pmid: [15215380](https://pubmed.ncbi.nlm.nih.gov/15215380/)
- Y.-C. Ma et al., Regulation of motor neuron specification by phosphorylation of neurogenin 2. *Neuron* **58**, 65–77 (2008). doi: [10.1016/j.neuron.2008.01.037](https://doi.org/10.1016/j.neuron.2008.01.037); pmid: [18400164](https://pubmed.ncbi.nlm.nih.gov/18400164/)
- G. S. Richards et al., Sponge genes provide new insight into the evolutionary origin of the neurogenic circuit. *Curr. Biol.* **18**, 1156–1161 (2008). doi: [10.1016/j.cub.2008.06.074](https://doi.org/10.1016/j.cub.2008.06.074); pmid: [18674909](https://pubmed.ncbi.nlm.nih.gov/18674909/)
- I. Dror, R. Rohs, Y. Mandel-Gutfreund, How motif environment influences transcription factor search dynamics: Finding a needle in a haystack. *BioEssays* **38**, 605–612 (2016). doi: [10.1002/bies.201600005](https://doi.org/10.1002/bies.201600005); pmid: [27192961](https://pubmed.ncbi.nlm.nih.gov/27192961/)
- S. Inukai, K. H. Kock, M. L. Bulyk, Transcription factor-DNA binding: Beyond binding site motifs. *Curr. Opin. Genet. Dev.* **43**, 110–119 (2017). doi: [10.1016/j.gde.2017.02.007](https://doi.org/10.1016/j.gde.2017.02.007); pmid: [28359978](https://pubmed.ncbi.nlm.nih.gov/28359978/)
- M. Slattery et al., Absence of a simple code: How transcription factors read the genome. *Trends Biochem. Sci.* **39**, 381–399 (2014). doi: [10.1016/j.tibs.2014.07.002](https://doi.org/10.1016/j.tibs.2014.07.002); pmid: [25129887](https://pubmed.ncbi.nlm.nih.gov/25129887/)
- D. A. Cusanovich et al., The cis-regulatory dynamics of embryonic development at single-cell resolution. *Nature* **555**, 538–542 (2018). doi: [10.1038/nature25981](https://doi.org/10.1038/nature25981); pmid: [29539636](https://pubmed.ncbi.nlm.nih.gov/29539636/)
- N. Nakanishi, S. Sogabe, B. M. Degnan, Evolutionary origin of gastrulation: Insights from sponge development. *BMC Biol.* **12**, 26 (2014). doi: [10.1186/1741-7007-12-26](https://doi.org/10.1186/1741-7007-12-26); pmid: [24678663](https://pubmed.ncbi.nlm.nih.gov/24678663/)
- I. S. Peter, E. H. Davidson, Evolution of gene regulatory networks controlling body plan development. *Cell* **144**, 970–985 (2011). doi: [10.1016/j.cell.2011.02.017](https://doi.org/10.1016/j.cell.2011.02.017); pmid: [2144487](https://pubmed.ncbi.nlm.nih.gov/2144487/)
- ENCODE Project Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012). doi: [10.1038/nature11247](https://doi.org/10.1038/nature11247); pmid: [22955616](https://pubmed.ncbi.nlm.nih.gov/22955616/)

49. R. E. Thurman *et al.*, The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82 (2012). doi: [10.1038/nature11232](https://doi.org/10.1038/nature11232); pmid: [22955617](https://pubmed.ncbi.nlm.nih.gov/22955617/)
50. E. Z. Kvon *et al.*, Genome-scale functional characterization of *Drosophila* developmental enhancers in vivo. *Nature* **512**, 91–95 (2014). doi: [10.1038/nature13395](https://doi.org/10.1038/nature13395); pmid: [24896182](https://pubmed.ncbi.nlm.nih.gov/24896182/)
51. J. F. Schmitz, F. Zimmer, E. Bornberg-Bauer, Mechanisms of transcription factor evolution in Metazoa. *Nucleic Acids Res.* **44**, 6287–6297 (2016). doi: [10.1093/nar/gkw492](https://doi.org/10.1093/nar/gkw492); pmid: [27288445](https://pubmed.ncbi.nlm.nih.gov/27288445/)
52. C. Lee, C.-H. Huang, LASAGNA: A novel algorithm for transcription factor binding site alignment. *BMC Bioinformatics* **14**, 108 (2013). doi: [10.1186/1471-2105-14-108](https://doi.org/10.1186/1471-2105-14-108); pmid: [23522376](https://pubmed.ncbi.nlm.nih.gov/23522376/)
53. E. Wingender, T. Schoeps, M. Haubrock, M. Krull, J. Dönitz, TFClass: Expanding the classification of human transcription factors to their mammalian orthologs. *Nucleic Acids Res.* **46**, D343–D347 (2018). doi: [10.1093/nar/gkx987](https://doi.org/10.1093/nar/gkx987); pmid: [29087517](https://pubmed.ncbi.nlm.nih.gov/29087517/)

ACKNOWLEDGMENTS

We thank L. Pennacchio for helpful discussions; G. Richards, I. Chang, E. Frampton, W. Luu, D. Garfield, S. Barreda-Pitcairn, and

the UQ IMB high-performance computing team for technical support; and the Victor Chang Cardiac Research Institute Innovation Centre, funded by the New South Wales Government.

Funding: Supported by Australian Research Council (ARC) awards DE160100755 and DP200100250 and University of Queensland Early Career Grant UQECR1832697 (E.S.W.); National Health and Medical Research Council (NHMRC) grants APP1107643 and APP116400 and fellowship APP1111169 and ARC grant DP200100250 (M.F.); and ARC grants DP160100573 and FL110100044 (B.M.D.). The Australian Regenerative Medicine Institute is supported by grants from the State Government of Victoria and the Australian Government. **Author contributions:** E.S.W., M.F., and B.M.D. were involved in the design of the experiments; E.S.W., D.Z., S.Z.T., N.L.B., and V.G. performed experiments; E.S.W., S.Z.T., D.Z., M.F., B.M.D., N.L.B., B.M.H., E.S., F.G., G.V., K.K., and E.M. were involved in data analyses; E.S.W., M.F., and B.M.D. wrote the manuscript; and E.S.W., S.Z.T., D.Z., and M.F. were involved in figure construction. All authors commented on the manuscript. **Competing interests:** The authors do not have any competing interests. **Data and materials**

availability: Sequencing data generated in this study can be found on ArrayExpress under E-MTAB-7846. Other data generated by this work are available in the supplementary materials. Code is available at Zenodo under DOI: [10.5281/zenodo.4001143](https://doi.org/10.5281/zenodo.4001143) and also

at <https://github.com/ewonglab/bagofmotifs>. The enhancer reporter transgenic zebrafish lines as well as corresponding constructs to establish the transgenic animals and *Amphimedon queenslandica* tissue samples are available from M.F., E.S.W., and B.M.D. under a material transfer agreement with the Centenary Institute, the Victor Chang Cardiac Research Institute, or the University of Queensland, respectively.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6517/eaax8137/suppl/DC1

Materials and Methods

Supplementary Text

Figs. S1 to S25

Tables S1 to S19

References (54–92)

Supplementary Code

Weeder Matrix and Output

MDAR Reproducibility Checklist

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

26 April 2019; resubmitted 29 April 2020

Accepted 30 September 2020

[10.1126/science.aax8137](https://doi.org/10.1126/science.aax8137)

RESEARCH ARTICLE SUMMARY

CRIMINAL JUSTICE

Behavioral nudges reduce failure to appear for court

Alissa Fishbane, Aurelie Ouss*, Anuj K. Shah

INTRODUCTION: Each year, millions of people in the United States are required to appear in court for low-level offenses. Many defendants miss their court dates. Criminal justice policy often uses punitive sanctions to deter these failures to appear. For example, when defendants fail to appear, arrest warrants are issued, which draws defendants further into the criminal justice system. These policies presume that defendants pay attention to these penalties and weigh them when deciding whether to appear in court. In this study, we explore a different possibility for why defendants might miss court—simple human error. Although defendants are given all of the relevant information they need, they might be insufficiently aware of this information—it might not be salient enough or defendants might forget it.

RATIONALE: We conducted two large-scale field studies in New York City evaluating interventions to make defendants more aware of court

information. In the first study, we redesigned the summons form that defendants receive for low-level offenses. The old summons form listed court information at the bottom of the form below less important information (e.g., the defendant's physical characteristics and details about the issuing officer), which makes it easy to overlook. The redesigned form moves court information to the top of the ticket, where people are more likely to see it. It also clearly states in bold typeface on the front of the form that missing the assigned court date will lead to a warrant. In the second study, we augmented the redesigned form by sending text messages to highlight critical court information for defendants in the week leading up to their court appearance.

In laboratory experiments, we tested how quickly people could identify, and how well they could remember, court information on the old and new forms. Then, we examined laypeople's and experts' beliefs about whether

failures to appear were intentional and how these beliefs affected their support for nudges to reduce failures to appear.

RESULTS: The redesigned summons form and text messages reduced failures to appear on average by 13 and 21%, respectively. In our laboratory experiments, people who saw the new forms identified court information more quickly (and recalled it more accurately). This suggests that a meaningful proportion of defendants who fail to appear are not intentionally skipping court but are effectively unaware of court.

We estimate that these nudges helped avoid at least 30,000 arrest warrants being issued over 3 years, and they resulted in approximately 20,000 people having their cases fully dismissed instead of having an open warrant. We see suggestive evidence that these nudges were more effective for defendants living in poorer neighborhoods. Additionally, the bulk of warrants avoided are for defendants living in poorer neighborhoods and neighborhoods with higher proportions of Black and Hispanic residents—though this reflects, rather than repairs, existing disparities in how summonses are issued.

However, in our laboratory experiments on people's attitudes, we find that laypeople believe failures to appear for court are more intentional than other failures to act (e.g., missing medical appointments). And although laypeople can be prompted to see how defendants might mistakenly miss court, their default assumption is to see it as intentional. This assumption reduces support for these nudges and increases support for more punitive approaches to reducing failures to appear. Criminal justice experts (e.g., prosecutors and judges), however, do not share this intuition—they are more likely to believe that failures to appear are unintentional and to support these nudges.

CONCLUSION: This work suggests that there is a straightforward explanation for why many defendants miss court—information about their court dates is not sufficiently salient. Although simple nudges can remedy this, policy-makers have been slow to experiment with and adopt them. Instead, criminal justice policy on failures to appear often hews closer to laypeople's intuitions. Widespread adoption of interventions such as these might depend on a shift in the assumptions of why failures to appear happen. Otherwise, these policies risk merely punishing people, not deterring actions. ■

SUMMONS PROCESS



Police officer issues court summons



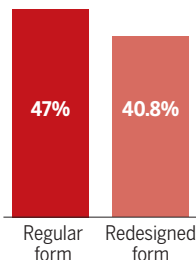
On court date, defendant shows up for hearing

INTERVENTION #1

REDESIGN CITY'S SUMMONS FORM

- Label as court appearance ticket
- Move court date and time near top
- Clearly indicate court location
- Highlight consequences of missing court

Failure-to-appear rates



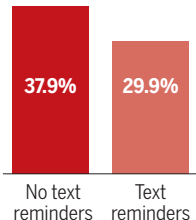
23,000 Warrants avoided

INTERVENTION #2

SEND TEXT REMINDERS IN ADVANCE OF DEFENDANT'S COURT DATE



Failure-to-appear rates



7,800 Warrants avoided



Defendant doesn't show up and court issues arrest warrant

Reducing failures to appear. In two field studies, we made court information more salient by redesigning the court summons form and sending text messages to defendants. We find that these interventions significantly reduced the rate at which defendants missed their court dates for low-level offenses, and fewer arrest warrants were issued as a result.

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

*Corresponding author. Email: aouss@sas.upenn.edu
Cite this article as A. Fishbane et al., *Science* 370, eabb6591 (2020). DOI: 10.1126/science.abb6591

S READ THE FULL ARTICLE AT
<https://doi.org/10.1126/science.abb6591>

RESEARCH ARTICLE

CRIMINAL JUSTICE

Behavioral nudges reduce failure to appear for court

Alissa Fishbane¹, Aurelie Ouss^{2*}, Anuj K. Shah³

Each year, millions of Americans fail to appear in court for low-level offenses, and warrants are then issued for their arrest. In two field studies in New York City, we make critical information salient by redesigning the summons form and providing text message reminders. These interventions reduce failures to appear by 13 to 21% and lead to 30,000 fewer arrest warrants over a 3-year period. In laboratory experiments, we find that whereas criminal justice professionals see failures to appear as relatively unintentional, laypeople believe they are more intentional. These lay beliefs reduce support for policies that make court information salient and increase support for punishment. Our findings suggest that criminal justice policies can be made more effective and humane by anticipating human error in unintentional offenses.

More than 10 million people are arrested each year in the United States, and millions more are issued summonses that draw them into the criminal justice system and require court appearances. Most of these arrests and summonses are for low-level offenses (1), and many of these defendants end up missing their court dates. In New York City alone, for example, we calculate that ~40% of defendants (or 100,000 people) missed their court date for low-level offenses in 2015. These failures to appear for court add to the original offense—defendants are held in contempt of court, and an arrest warrant is issued, which is supposed to act as a deterrent. Failures to appear are also common for more serious offenses, such as felonies and misdemeanors. In those contexts, the consequences can be even more costly. Partly to reduce the risk of failures to appear, judges often assign felony and misdemeanor defendants to pretrial detention, which contributes to the scale of incarceration in the United States, where 500,000 people are in jail awaiting their trials each day. Other felony and misdemeanor defendants are asked to post monetary bail, which acts as collateral to incentivize appearance in court. These deterrence policies aim to reduce failures to appear by increasing the penalties associated with them. These policies will be most effective if defendants pay attention to the penalties and make an intentional decision whether to skip court or not on the basis of the costs and benefits of doing so.

In this paper, we explore a different possibility for why defendants might miss court—simple human error. Although defendants are

given all of the relevant information they need (e.g., when and where to appear for court and what the consequences are for missing court), they might be insufficiently aware of this information. This could happen for various reasons: The information might not be salient enough, or defendants might simply forget it as their court date approaches. In other words, many failures to appear may occur not because defendants are intentionally showing contempt of court, but rather because existing policies do not allow enough room for error.

In some ways, this hypothesis is fairly straightforward. Insufficient awareness can explain various other failures to act—for example, failures to save money or pay bills (2, 3), failures to get immunizations (4–6), and even student failures to matriculate in college (7). In criminal justice, raising the awareness of consequences can reduce misconduct (8). However, the simplicity of this hypothesis makes it all the more startling in this context. A single failure to act in other domains might have few direct consequences, and those consequences might be delayed (often by years). By contrast, failures to appear for court have direct, immediate, and severe consequences—conditions that criminology suggests should reduce misconduct (9). It might seem that court dates and the threat of arrest warrants would not be things that people would simply forget or overlook.

If our hypothesis is true, however, then policies that focus only on punishment may be poorly targeted for reducing failures to appear. It might in fact be more cost-effective, and more humane, to make court information salient for defendants.

We find evidence for this from two large-scale field studies conducted in New York City. In the first study, we redesign court summons forms to simplify how information is presented. In the second study, we augment the redesigned form by sending text messages that highlight critical court information for defen-

dants. These interventions reduce failures to appear, on average, by 13 and 21%, respectively, which suggests that a meaningful proportion of defendants who fail to appear are not intentionally skipping court but are effectively unaware of court. In a series of laboratory experiments, we find evidence in support of this hypothesis, as the redesigned summons form improved participants' identification and recall of court information.

However, we also find that laypeople's intuitions about why failures to appear happen might lead them to overlook the value of these interventions. Specifically, laypeople believe that failures to appear are relatively intentional, and this belief leads to lower popular support for interventions like the ones we tested here. When prompted, laypeople can appreciate how human error might play a role in failures to appear, and this increases their support for the interventions we tested. However, laypeople's default intuition is that failures to appear stem from intentional decisions to skip court. Notably, criminal justice experts (e.g., prosecutors and defense attorneys) are more likely to believe that failures to appear are unintentional and are more likely to support these interventions. Nevertheless, current criminal justice policy often aligns more with the intuitions of our samples of laypeople rather than those of experts.

Field studies on nudges for defendants

In our study, we focus on criminal summonses, which are typically issued for the lowest level of criminal offenses, in New York City. Criminal summonses typically result from quality-of-life offenses, such as open containers, disorderly conduct, or park trespassing (see table S1 for more information on summonses and descriptive statistics on summons recipients). In 2015, the most recent year before our study period, 256,488 summonses were issued that required court appearances [though summonses have since declined (10)]. For these offenses, defendants are typically not arrested, taken into custody, detained pretrial, or required to post bail. Instead, they are given a summons form and are required to appear in court 60 to 90 days later, with some flexibility the week before the scheduled court date. However, if they fail to appear in court, a warrant is opened for their arrest, which means that future interactions with the police are more likely to result in an arrest, possibly even after an illegal stop. Failure to appear is also a separate violation that can carry a fine of \$250 and a penalty of up to 15 days in jail. If defendants voluntarily show up to court at a later date, the warrant will often be vacated. However, many defendants may effectively be unaware that they have open warrants. Historically, failure-to-appear rates are ~40% for summonses that require a court appearance.

¹Idea42, New York, NY 10004, USA. ²Department of Criminology, University of Pennsylvania, Philadelphia, PA 19104, USA. ³Booth School of Business, University of Chicago, Chicago, IL 60637, USA.

*Corresponding author. Email: aouss@sas.upenn.edu

A

CRC-3206 (5/12)

Complaint/Information
The People of the State of New York vs.

Name (Last, First, MI)
Street Address
City
State
Zip Code
Apt. No.

ID/License Number
State
Type/Class
Expires (mm/dd/yy)
Sex

Date of Birth (mm/dd/yy)
Hi
Wt
Eyes
Hair
Plate/Reg

Reg State
Expires (mm/dd/yy)
Plate Type
Veh Type
Make
Year
Color

The Person Described Above is Charged as Follows:

Time 24 Hour (hh:mm)
Date of Offense (mm/dd/yy)
County

Place of Occurrence
NY State
Precinct

In Violation of Section
Subsection
VTL
Admin Code
Penal Law
Park Rules
Other

Title of Offense:

Bronx Criminal Court - 215 E 161st Street, Bronx, NY 10451
Kings Criminal Court - 346 Broadway, New York, NY 10013
Redhook Community Justice Center - 88-94 Visitation Place, Brooklyn, NY 11231
New York Criminal Court - 346 Broadway, New York, NY 10013
Midtown Community Court - 314 W 54th Street, New York, NY 10019
Queens Criminal Court - 120-55 Queens Boulevard, Kew Gardens, NY 11415
Richmond Criminal Court - 67 Targee Street, Staten Island, NY 10304

Defendant stated in my presence (in substance):

I personally observed the commission of the offense charged herein. False statements made herein are punishable as a Class A Misdemeanor pursuant to section 210.45 of the Penal Law. Affirmed under penalty of law.

Complainant's Full Name Printed
Rank/Full Signature of Complainant
Date Affirmed (mm/dd/yy)

Agency
Tax Registry #
Command Code

The person described above is summoned to appear at NYC Criminal Court
Summons Part
County

Date of Appearance (mm/dd/yy)
At 9:30 a.m.

DEFENDANT'S COPY

B

CRC-3206 (7/15)

Criminal Court Appearance Ticket

Name (Last, First, MI)
Date of Birth (mm/dd/yy)

Cell Phone Number (where court may contact you)
Home Phone Number (where court may contact you)

Show up to court on:
Court Appearance Date (mm/dd/yy):
at 9:30 a.m.

Your court appearance location: ☐ Other (specify) _____

☐ Bronx Criminal Court
☐ Kings & New York Criminal Court
☐ Midtown Community Court
☐ Redhook Community Justice Center
☐ Queens Criminal Court
☐ Richmond Criminal Court

****To avoid a warrant for your arrest, you must show up to court.****
At court, you may plead guilty or not guilty.
Please see back for exceptions for Public Consumption of Alcohol and Public Urination offenses.

Court Locations: You must appear at the court location identified above.

Bronx Criminal Court 215 E 161st Street, Bronx, NY 10451
Kings & New York Criminal Court 346 Broadway, New York, NY 10013
Redhook Community Justice Center 88-94 Visitation Place, Brooklyn, NY 11231
Midtown Community Court 314 W 54th Street, New York, NY 10019
Queens Criminal Court 120-55 Queens Boulevard, Kew Gardens, NY 11415
Richmond Criminal Court 26 Central Ave, Staten Island, NY 10301

You are Charged as Follows:

Title of Offense:
Time 24 Hour (hh:mm)
Date of Offense (mm/dd/yy)
County

Place of Occurrence
Precinct

In Violation of Section
Subsection
VTL
Admin Code
Penal Law
Park Rules
Other

For Additional Information and Questions:

Visit the website or call the number below for additional information about your court appearance and translation of this document.

www.mysummons.nyc
OR
Call 646-760-3010

Defendant stated in my presence (in substance):

I personally observed the commission of the offense charged herein. False statements made herein are punishable as a Class A Misdemeanor pursuant to section 210.45 of the Penal Law. Affirmed under penalty of law.

Complainant's Full Name Printed
Rank/Full Signature of Complainant
Date Affirmed (mm/dd/yy)

Agency
Tax Registry #
Command Code

DEFENDANT'S COPY

Note - If you do not follow these instructions a warrant may be ordered for your arrest.
To Plead Not Guilty. You must come to court at the place and time shown on the front of this summons.
To Plead Guilty. You must come to court at the place and time shown on the front of this summons. However, if you are charged with **Public Consumption of Alcohol**, or **Public Urination**, you may plead guilty by mail. If you plead guilty by mail you do not have to come to court.
To Plead Guilty by Mail within 10 days of the date this summons was issued (To one of the above two charges only):
* Complete and sign the "Guilty Plea by Mail Form" below.
* Prepare a check or money order payable to NYC Criminal Court for **\$25 for Public Consumption of Alcohol** or **\$50 for Public Urination**.
* Write the summons number on the check or money order.
DO NOT SEND CASH.
* Mail the check or money order with this completed form to:
NYC CRIMINAL COURT
P.O. BOX 555
NEW YORK, N.Y. 10031-0555
The court record of this case will show your guilty plea to an offense that is not a crime under New York law. This guilty plea will be a public record available at the courthouse. Consult with an attorney if you have questions about this.

Guilty Plea by Mail Form
For Public Consumption of Alcohol or Public Urination Only
I hereby plead guilty to the following charge: (Place a mark next to the correct charge.)
Public Consumption of Alcohol [Administrative Code section 10-125 (b)], a violation not a crime as charged on this summons.
* I understand that this violation is punishable by a fine of not more than \$25 or imprisonment of up to five days or both.
* I agree that the sentence imposed will be a \$25 fine.
Public Urination [Administrative Code section 16-118(6)], a violation not a crime as charged on this summons.
* I understand that this violation is punishable by a fine of \$50-\$250 or imprisonment for up to ten days or both.
* I agree that the sentence imposed will be a \$50 fine.
By entering a plea of guilty to this charge I agree to the following:
* I waive arraignment in open court, the right to the assistance of a lawyer, and the right to receive a copy of the accusatory instrument.
* I understand that a plea of guilty to the charge is equivalent to a conviction after a court trial.
* I understand that the court can refuse my guilty plea. If this occurs my payment will be returned and the court will notify me in writing when and where to come to court regarding this summons.
I plead guilty to and admit committing the offense as set forth in this summons.
Signature _____ Date _____
Check or Money Order Payable to NYC Criminal Court Must Be Enclosed.

Fig. 1. Old and redesigned New York City summons forms. (A) Front and back of previous version of the New York City summons form. **(B)** Front and back of the redesigned version of the form.

For our analyses, we rely on administrative data collected by the New York State Office of Court Administration. The data contain defendant gender, date of birth, and address; information about the violation; and court outcomes (see supplementary materials for more details). Our primary sample includes all 323,922 summonses issued in New York City between 1 January 2016 and 14 June 2017.

At the time of our first study, the only way that defendants were notified of their court dates was on the summons ticket they received at the time of the original offense. The entire policy to inform defendants about their court date and deter them from skipping court depended on this form. However, the form's design prioritized information about the original offense rather than information about the defendant's court appearance. For example, it devoted substantial space to describing the defendant (e.g., height, weight, hair color, and the defendant's vehicle) and the violation. The defendant's court date was written at the bottom of the form, below details about the officer issuing the summons form. Only on the back of the form was it mentioned that arrest warrants are issued for those who fail to comply. Given that this information was so easy to overlook, many defendants might have been insufficiently aware of when they were expected to appear in court and what the consequences were for missing court.

We worked with the Mayor's Office of Criminal Justice, New York City Police Department, and the New York State Office of Court Administration to redesign the summons ticket to make relevant information more salient. The new form prominently features the appearance date and court location at the top of the ticket, where people are more likely to see it. It also clearly states in bold typeface on the front of the form that missing the assigned court date will lead to a warrant (see Fig. 1 for old and new forms). If this form reduces failures to appear, that suggests that many defendants might have missed court simply because they were unaware of important information.

We evaluated the effectiveness of this intervention using a regression discontinuity design, which compared failure-to-appear rates for defendants who happened to be among the first to receive a new form given by a particular police officer versus failure-to-appear rates of defendants who were among the last to receive the old form given by that same officer. New forms were gradually rolled out between

March and August of 2016. Police officers only had one pad of summons forms with them. They switched from old forms to new forms when they used up their pad of old forms or when their supervisor handed out pads of new forms. We obtained the identification number of the issuing officer for each form. Every summons form has a serial number, and we can match serial numbers to old or new forms. We construct an officer's switch date by randomly choosing a date between when that officer issues their last old form and their first new form. By construction, before an officer's switch date, all of the forms issued were of the old design; but after the switch date, 97.6% of forms were of the new design, which suggests very good compliance (fig. S1).

Because the introduction of the new forms was staggered, we can also control for seasonality (which is important because types of crimes and failures to appear both vary seasonally) and other time trends. Our main identifying assumption is that there are no observable differences in defendants' characteristics on the basis of whether they received a summons form before or after an officer's switch date. We in fact find no differences in prior summonses or failures to appear, types of offenses, or their predicted failure-to-appear likelihood based on observables, and we observe only a small difference in gender (figs. S2 and S3 and table S2). Because the only notable difference before and after the switch date is the kind of form issued, any difference in failure-to-appear rates for defendants on either side of the switch date can plausibly be attributed to the redesigned form (see supplementary materials for details and robustness checks regarding our identification strategy).

Our main results are presented in Fig. 2 and table S3. Figure 2 presents failure-to-appear

rates for defendants issued forms just before and just after officers' switch dates. Failure-to-appear rates are lower just after the introduction of the new summons forms. To estimate the magnitude of this drop in failure-to-appear rates, we follow the approach of Calonico *et al.* (11, 12), which allows us to obtain consistent estimates when we include covariates. We find that the new forms reduced failures to appear by 6.2 percentage points, or by 13.2% relative to the 47% baseline failure-to-appear rate in the estimation bandwidth ($P < 0.001$). Note that in Fig. 2, the failure-to-appear rate appears to be increasing before the switch date. This is because most officers switched to the new form between May and July, when failure-to-appear rates are highest. We observe this seasonality of failure-to-appear rates in other years, but the drop in failure-to-appear rates after the switch date appears to be the specific result of introducing the redesigned forms (see supplementary materials discussion of robustness checks and fig. S5).

Our second field study provides more direct evidence that failures to appear might stem from defendants' lack of awareness of critical information. In this study, we tested whether failures to appear could be further reduced by texting defendants their court information (date and location) and information about the consequences of missing court. Defendants who received the new summons form could provide their cell phone number to the citing officer, though doing so was not mandatory. All defendants who provided their cell phone number were included in this evaluation.

Approximately 11% of defendants (23,243) provided their phone numbers. There appears to be positive selection in who provided a phone number. For example, the failure-to-appear rate of people who provided a phone

number and were randomized to the control group is 37.9%, relative to 40.8% for defendants who did not provide a phone number ($P < 0.001$). Still, failures to appear were very frequent even among people who provided a phone number (see supplementary materials for a more detailed discussion of and tests for external validity on the basis of this sample).

Defendants were randomly assigned to one of four conditions. The control group received no text messages. In the other groups, summons recipients received three messages: a message 7 days before, a message 3 days before, and a message 1 day before their scheduled court date. We varied the content of the messages to better identify what information is most effective at reducing failures to appear. In the consequences group, defendants received messages that described their court date and location and also told them that a warrant would be opened and they might be arrested if they missed their court date. In the plan-making group, defendants received messages that described their court date and location and also prompted them to make a plan to attend court, including marking their calendars, setting an alarm, and looking up directions (but there was no mention of consequences). In the combination group, defendants received a mix of the messages from the consequences and plan-making groups. Analyses below were preregistered. Defendants who missed court were also randomized to receive different follow-up messages, but these were not part of our primary analyses in our preanalysis plan and are not discussed here.

We can evaluate the effectiveness of these messages in a few ways. First, do any text messages reduce failures to appear? As shown in table S5, relative to a 37.9% failure-to-appear rate in the control group, receiving any text message reduced failures to appear by 8 percentage points, which represents a 21% relative reduction ($P < 0.001$). Second, the differences across treatment groups also provide some evidence for why these messages are effective (Fig. 3 and table S5). The consequences and combination messages were most effective, reducing failures to appear by 8.9 and 9.9 percentage points relative to the control group (23.5 and 26.1% relative reductions; $P < 0.001$ in both cases), respectively. This suggests that a substantial proportion of defendants miss court because they are unaware of the consequences. The plan-making messages, which did not mention the consequences of failure to appear, also significantly reduced failures to appear by 6 percentage points (15.8%; $P < 0.001$). These results build on an earlier smaller-scale study that examined the effectiveness of postcard reminders for defendants in a context where baseline failure-to-appear rates were substantially lower than the current context (13).

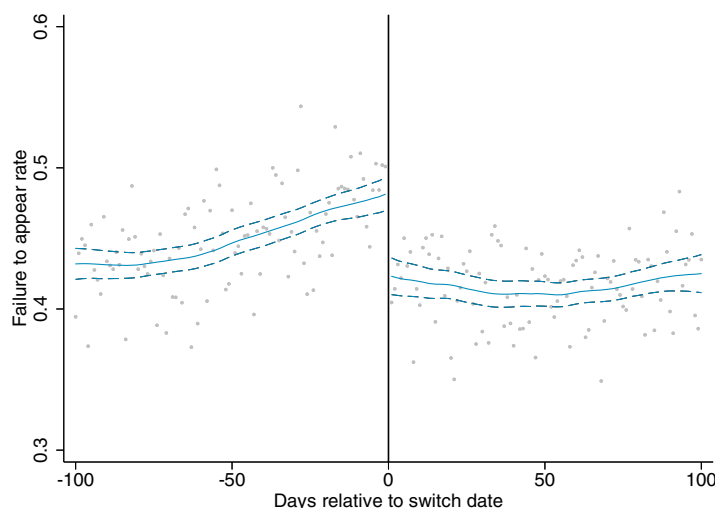
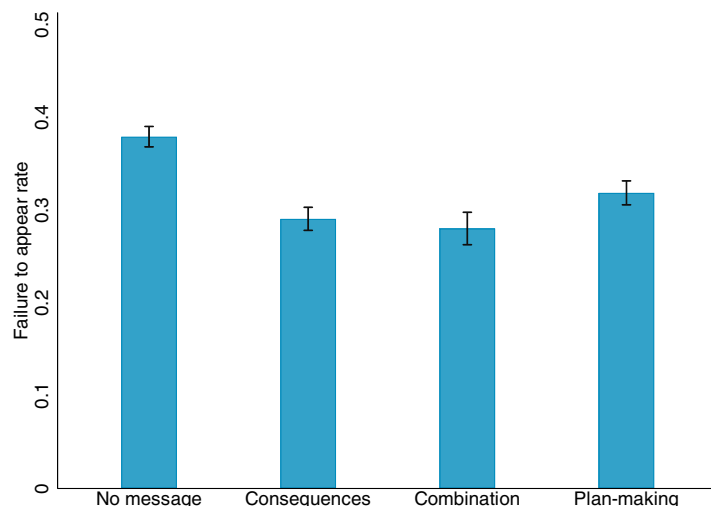


Fig. 2. Scatterplot of daily averages of failure-to-appear rate relative to the switch date from old to new forms. Solid lines are local-polynomial regression lines, and dashed lines represent the 95% confidence intervals.

Fig. 3. Failure-to-appear rate by text message treatment arm. Errors bars represent 95% confidence intervals.



The fact that these reminders are effective suggests that a meaningful proportion of defendants missed court because they lacked the most basic information about their scheduled appearance. It is possible that the plan-making component of the messages might have helped defendants show up to court, but even the simple consequences message—which just contained information about their court appearance—reduced failures to appear.

Laboratory experiments on mechanisms and punitive attitudes

It is possible that our interventions were effective for other reasons besides just making defendants more aware of court information. For example, perhaps defendants were already aware of the information—they noticed it and remembered it—but our interventions led them to believe punishment for failure to appear was more likely. Below, we find stronger evidence for the awareness hypothesis in two laboratory experiments. We then explore whether people might underappreciate how insufficient awareness can lead to failures to appear and whether they might overlook the value of the nudges we tested.

In our first two laboratory experiments, we tested whether the redesigned forms increased awareness of court information (see Materials and methods and the supplementary materials for further details on all laboratory experiments). In the first laboratory experiment, 232 participants from Amazon Mechanical Turk (MTurk) first read background information about failures to appear in New York City. Participants were then shown a summons form and asked to identify three pieces of information on the form: the defendant's court date and time, the defendant's court location, and the defendant's alleged offense. Participants were randomly assigned to see either the old form or new form, and they clicked on the parts of the form that contained the information. The

new form simply moves court information to the top but leaves the position of information about the alleged offense unchanged. We recorded how long (in milliseconds) it took participants to identify each piece of information. We expected that people would be faster to identify court information in the new form (because this information was prioritized at the top of the form) but would not be any faster to identify information about the alleged offense (because its position was unchanged).

Participants who saw the new form identified the court date and time more quickly ($\text{Mean}_{\log(\text{reaction time})} = 4.46$; $\text{SD} = 0.46$) than participants who saw the old form ($M_{\log(\text{RT})} = 4.72$; $\text{SD} = 0.37$), per a t test with unequal variances [$t(229.17) = 4.86$; $P < 0.001$]. This was also true for identifying the court location (new form: $M_{\log(\text{RT})} = 4.59$; $\text{SD} = 0.40$ versus old form: $M_{\log(\text{RT})} = 4.70$; $\text{SD} = 0.35$) [$t(230) = 2.37$; $P = 0.02$]. Participants who saw the old and new forms did not significantly differ in how quickly they identified the alleged offense [$t(230) = 0.39$; $P = 0.69$]. Thus, it is clear that people more easily identify information at the top of the form, and moving court information there makes it more accessible.

Laboratory experiment 2 extends these results by testing whether the new form actually improves recall of court information. We recruited 725 New York City residents on MTurk. Participants were told to imagine receiving a summons form for disorderly conduct. They were then randomized to see either the old summons form or the new summons form. They completed a brief filler task (to create a gap before subsequent questions) and then responded to several questions. Of primary interest were the questions where they were asked to recall the penalty for failure to appear, the court date, and the court location.

First, we found that participants who saw the new forms were more likely to correctly recall their court date (new form, 38%; old form,

19%; $P < 0.001$) and court location (new form, 46%; old form, 26%; $P < 0.001$). Moreover, we found that participants who saw the new forms were more likely to correctly recall that the penalty for failure to appear was a warrant (new form, 52%; old form, 41%; $P = 0.003$; see table S7). The results are similar with and without covariates, including whether a person had received a summons in the past.

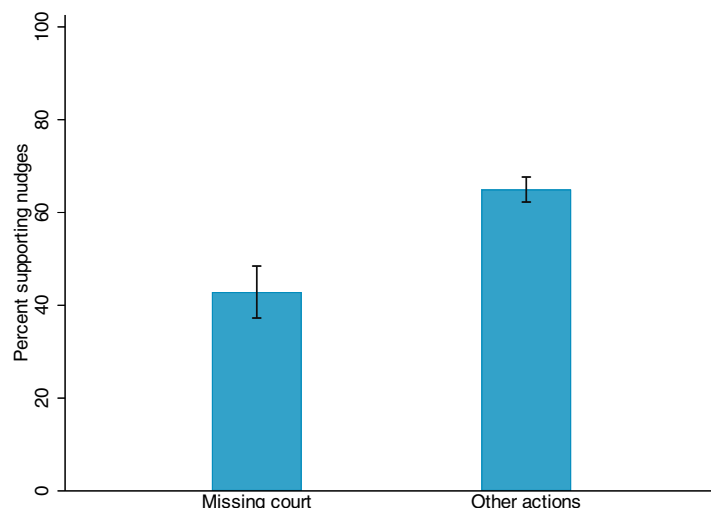
It is worth noting that, in some ways, this experiment makes it fairly easy to remember the court information. Participants were asked about the information shortly after seeing it. Nevertheless, recall rates are lower for the old forms. This suggests that defendants who received the old form might have been unaware of court information simply because it was not communicated effectively. Participants also rated the forms on other dimensions, such as whether they made participants feel angry or confused and whether participants felt the tickets were fair or reasonable. We do not see any reliable differences in ratings across these dimensions. The main difference appears to be that the new form made it easier to find information about court and the consequences of missing court (table S8).

Given how straightforward these interventions are, why might they have only recently (and not yet widely) been implemented? In our remaining laboratory experiments, we consider whether people's mental models of criminal justice might lead them to underestimate the effectiveness of interventions like these.

In laboratory experiment 3, participants ($N = 301$, MTurk sample) read five scenarios about people failing to take a required action: failing to appear for court, failing to pay an overdue bill, failing to show up for a doctor's appointment, failing to turn in paperwork for an educational program, and failing to complete a vehicle emissions test. These scenarios were selected to provide a cross section of different policy domains—criminal justice, personal finance, health, education, and environmental decisions—and because most are situations where reminders have proven effective. To limit differences across domains, participants were told that the following details applied to each scenario: the person was required to take an action in 60 days, the person did not want to take the action, there was a penalty for failing to act, and ultimately the person did not take the required action. For each scenario, participants rated how likely they thought it was that the person missed their appointment because they did not pay enough attention to the scheduled date or because they simply forgot. They also rated how likely it was that the person deliberately and intentionally decided to skip their appointment. Finally, participants were asked what they thought should be done to make sure that other people show up for their appointments, and they chose

Fig. 4. Participants' support for using nudges to reduce failures to appear in court and failures to complete other actions (from laboratory experiment 3).

Errors bars represent 95% confidence intervals.



one of three options: increase the penalty for failing to show up, send reminders to people about their appointments, or make sure that appointment dates are easy to notice on any paperwork. All scenarios were presented in a random order for each participant.

Relative to most other actions, participants rated failures to appear for court as less likely to be a result of forgetting ($M_{\text{court}} = 3.86$, $SD = 2.06$; $M_{\text{other actions}} = 4.24$, $SD = 1.45$; paired t test, $t(300) = 3.79$, $P < 0.001$) and more likely to be intentional ($M_{\text{court}} = 5.17$, $SD = 1.75$; $M_{\text{other actions}} = 4.82$, $SD = 1.29$; paired t test, $t(300) = 3.92$, $P < 0.001$). Next, we analyzed whether participants supported either of the nudges (sending reminders and making appointment information easy to notice) over stiffer penalties. Relative to all other actions, participants were least supportive of nudges to reduce failures to appear ($M_{\text{court}} = 43\%$, $SD = 50$; $M_{\text{other actions}} = 65\%$, $SD = 34$; paired t test, $t(300) = 8.13$, $P < 0.001$), as shown in Fig. 4 (see figs. S6 and S7 for breakdown by each action). It appears that people generally ascribe greater intentionality to failures to appear, and these intuitions may inform why people believe stiffer penalties are more effective than nudges for reducing failures to appear.

We explore this link further in laboratory experiment 4. Participants ($N = 304$, MTurk sample) read background information on summonses and failure-to-appear rates in New York City. Our main dependent variable was what participants thought should be done to reduce the failure-to-appear rate: increase the penalty for failing to show up, send reminders to people about their court dates, or make sure that court dates are easy to notice on the summonses. Participants were randomly assigned to one of three conditions. In the control condition, participants made their policy choice immediately after reading the background information. In the intentional condition, after

reading the background information, participants wrote down one reason why someone might purposely skip their court appearance, and then they made their policy choice. In the mistake condition, participants wrote down one reason why someone might accidentally miss their court appearance, and then they made their policy choice.

We find two notable results. First, participants' policy recommendations did not significantly differ between the control (63% supported nudges; i.e., reminders or making court dates easy to notice) and intentional (61%) conditions [$\chi^2(1, N = 304) = 0.09$; $P = 0.76$], which suggests that participants' default assumptions are that failures to appear are intentional. Second, 82% of participants supported nudges in the mistake condition, significantly more than in both the control [$\chi^2(1, N = 304) = 9.08$; $P = 0.003$] and the intentional conditions [$\chi^2(1, N = 304) = 10.53$; $P = 0.001$], which suggests that their attitudes are malleable, as shown in Fig. 5. Our participants are generally supportive of using nudges instead of stiffer punishments, and this is in line with previous work that has shown that people tend to hold favorable views of nudges (14, 15). Support for nudges in this experiment is also higher than what we found in laboratory experiment 3, perhaps because more background information (e.g., the baseline failure-to-appear rate) was provided to participants. Nevertheless, our data suggest that people's default assumption is that failures to appear are intentional, and this weakens support for nudges.

In laboratory experiment 5, we tested whether experts shared laypeople's intuitions. We recruited, through email listservs, a sample of criminal justice professionals (e.g., judges, prosecutors, and defense attorneys; for full recruitment details, demographics, and discussion of attrition, see supplementary materials). In total, 145 experts completed the full

study. The most common professions in our sample were prosecutor (58%) and defense attorney (17%).

There were two parts to the study. The first part was a direct replication of laboratory experiment 4. Experts' responses did not significantly vary across conditions ($P > 0.3$ in all conditions), but we found that the vast majority of experts (89% across conditions) favored using nudges over stiffer penalties, showing significantly greater support than that observed in our sample of laypeople [$\chi^2(1, N = 449) = 21.56$; $P < 0.001$].

In the second part of the study, participants were shown pictures of both the old and new form (they were not labeled as such), and they indicated whether they thought recipients of the old or new forms would be more likely to remember their court information and show up for their court appearance. A clear majority of experts thought that recipients of the new form would be more likely to remember their court date (86%) and court location (68%) and show up to court (69%). For comparison, we asked a sample of 301 MTurk participants these same questions. As shown in Fig. 5, experts thought that the new form would be more effective than laypeople, who showed no clear preference for the new form (court date, 49%; court location, 50%; show up to court, 47%; $P < 0.001$ in all cases).

It is possible that our sample of experts was particularly reform-minded relative to other experts. However, these results might suggest a notable tension. These experts seemed to view failures to appear as less intentional than laypeople, and they showed greater support for nudges to reduce failures to appear. This was true regardless of experts' jobs within the criminal justice system. Nevertheless, criminal justice policy on failures to appear seems to hew closer to laypeople's intuitions. It is common to try to deter failures to appear through the threat of punishment but rarer to use nudges to prevent them. These studies suggest that increasing the adoption of these nudges may depend partly on shifting policymakers' mental models of why offenses like failures to appear happen.

Effects across socioeconomic status and race

Failing to account for human error in the context of criminal justice policy has profound consequences, and these consequences are often borne by the poor and people of color, who are disproportionately affected by the criminal justice system. In exploratory analyses, we find some evidence of this disproportionate involvement in the context of failures to appear as well. We do not have reliable, individual-level data on summons recipients' wealth or race. However, we have data on home addresses, which we can match to census tract data. We can then use census

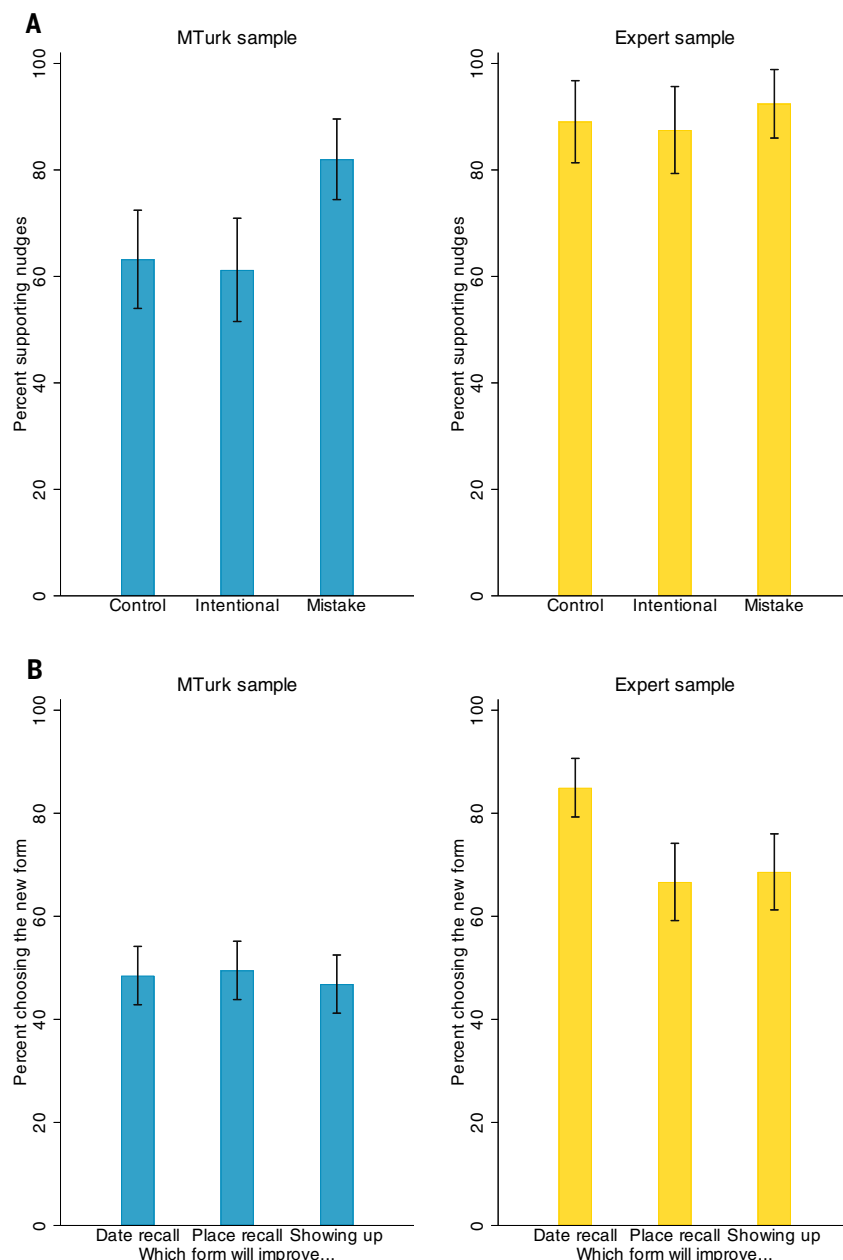


Fig. 5. Laypeople's and experts' beliefs about failures to appear and effectiveness of the redesigned summons forms. (A) Laypeople's (left) and criminal justice experts' (right) support for nudges to reduce failures to appear based on whether they thought about these failures as intentional or as a result of forgetting or were in the control condition. (B) Laypeople's (left) and criminal justice experts' (right) judgments of whether the redesigned summons form would improve recall of court information and the rate at which defendants showed up for court (from laboratory experiments 4 and 5). Error bars represent 95% confidence intervals.

tract-level data on poverty and racial composition as a proxy for defendants' wealth and race.

First, we ask whether the effectiveness of our interventions varied on the basis of the wealth and racial composition of defendants' neighborhoods (fig. S8 and tables S10 to S13). We see some suggestive evidence that the interventions were more effective for poorer defendants. Dividing our sample of summons recipients by quintiles, the treatment effect for

the text messaging intervention is 12.5 percentage points (27%) for defendants living in the bottom quintile of neighborhood wealth, which is significantly greater than the average of 6.4 percentage points (19%) for defendants living in other quintiles [treatment by poorest quintile interaction: $\beta = -0.058$, standard error (SE) = 0.017, $P = 0.001$]. The effectiveness of the form redesign does not significantly vary by wealth, although the results trend in the

same direction. The form redesign reduced failures to appear by 8 percentage points (15%) for defendants living in the bottom quintile of neighborhood wealth compared with a reduction of 5.7 percentage points (13%) for defendants living in other neighborhoods. These results are mixed, but it may be helpful to think about them in the context of recent work, which has suggested that poorer individuals must often cope with greater demands on their cognitive bandwidth (16–18). If poorer individuals are already dealing with greater cognitive demands, then our findings hint at the possibility that interventions such as those evaluated here might be particularly important for poorer defendants.

We do not find that the effectiveness of our interventions depends on the racial composition of defendants' neighborhoods. However, defendants who live in neighborhoods with a higher proportion of Black or Hispanic residents were less likely to give their phone numbers to officers. We cannot identify why this is, but it could reflect different policing practices (e.g., if officers do not ask for phone numbers as often in these neighborhoods) or mistrust between police and people of color (who may be reluctant to provide their phone numbers). If text message reminders are an effective way to reduce failures to appear (and thereby reduce open warrants), greater effort is needed to ensure that this intervention can benefit all communities.

Of course, nudges such as these are not sufficient to address larger, structural disparities in the criminal justice system. These dynamics are apparent when we examine whether wealth and racial composition of a neighborhood predicts the number of summonses issued (Fig. 6). We divided New York City census tracts into percentiles, where higher percentiles correspond to a greater proportion of residents living below the poverty line or a greater proportion of Black and Hispanic residents. We then regressed the number of summonses issued (per 1000 residents) on percentile. We find that more summonses are issued in poorer neighborhoods ($\beta = 0.29$, SE = 0.01, $P < 0.001$) and in neighborhoods with more Black and Hispanic residents ($\beta = 0.35$, SE = 0.01, $P < 0.001$). To put these differences in context, more than half of all summonses issued are in the poorest 30% of census tracts (where >18% of residents live below the poverty line). Similarly, more than half of all summonses issued are in the 32% of census tracts with the highest proportion of Black and Hispanic residents (where >80% of residents are Black or Hispanic). These differences in summonses issued are accompanied by failure-to-appear rates that are higher for defendants living in poorer neighborhoods (poorest quintile among summonses recipients: 53%, versus wealthiest quintile: 37%) and neighborhoods

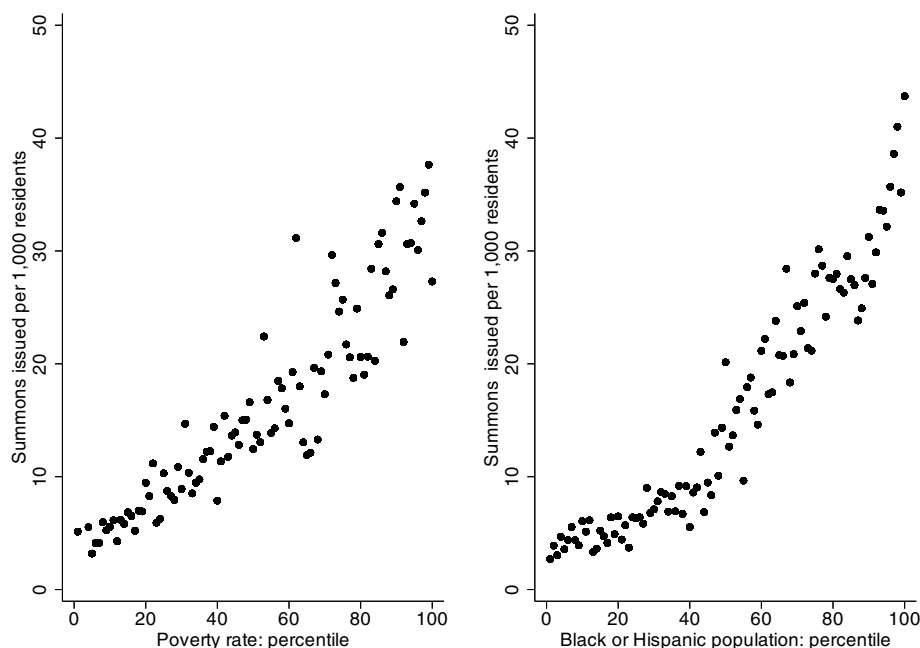


Fig. 6. Summonses, wealth, and race. Summonses issued by wealth (left) and racial composition (right) of census tracts. Census tracts are divided into percentiles, where higher percentiles correspond to a greater proportion of residents living below the poverty line or a greater proportion of Black and Hispanic residents.

with the highest proportion of Black and Hispanic residents (highest Black and Hispanic population quintile: 53%, versus lowest Black and Hispanic population quintile: 34%), which compounds the negative consequences of summonses on poor and minority citizens.

Given that summonses (and failures to appear) are disproportionately concentrated in neighborhoods that are poorer and have a higher proportion of Black and Hispanic residents, the benefits (i.e., reductions in failures to appear and open warrants) of implementing interventions such as ours may also be concentrated in these neighborhoods regardless of whether their effectiveness varies on the bases of wealth and race.

Discussion

This work suggests that there is a straightforward explanation for why many defendants miss court—information about their court dates is not sufficiently salient. However, policies have failed to fully account for this. Instead, these policies are often targeted toward reducing failures to appear as if they occur through intention rather than error.

By anticipating how human error can lead to failures to appear, our interventions have clear benefits for both defendants and the court system. We estimate that the form redesign and text message interventions have helped avoid at least 30,000 warrants being issued between August 2016 and September 2019 (see supplementary materials for detailed calculations). Moreover, ~66% of summonses

recipients see their case conditionally or unconditionally dismissed when they show up to court (19). This means that our interventions likely resulted in ~20,000 people having their cases fully dismissed instead of having an open warrant. It is also worth noting that the design of New York City's old summons form was not unique, as other major cities' forms have similar designs (for an example, see fig. S9). The benefits we see here are therefore likely to generalize to other cities as well.

Warrants can negatively affect many dimensions of people's lives, even if they do not result in an arrest. Sociologists have noted how warrants can have an effect of marking people, or designating their involvement in the criminal justice system, which both creates hassles for them and increases the chances of later escalation in criminal justice involvement (20). Ethnographic research has shown that people with open warrants often avoid places where they fear they could be identified and arrested. They may be less likely to go to hospitals for medical treatment or to show up to regular jobs, and they may frequently change their housing (21). They are also less likely to call the police to report crimes (even when they are the victims), and they are less likely to use social services and government assistance available to them, such as food stamps or job training programs (22). Open warrants can also be public record, which could negatively affect job prospects, housing, and a range of other outcomes for defendants. Though it is difficult to quantify these negative effects, it is clear

that by reducing warrants, these interventions can have cascading benefits.

Failures to appear are also costly to the criminal justice system itself. Court personnel time is wasted when defendants miss court, and efforts are diverted in issuing warrants. Each warrant costs approximately \$21 in judge and staff time (23), which translates into a savings of more than \$600,000 in court personnel time alone from these interventions. We can also estimate that each arrest would cost \$454 in police and court personnel time [based on (24); see supplementary materials for detailed calculations]. From our data, we cannot estimate how often warrants for summons failures to appear lead to arrests in New York City. However, a recent study in St. Louis found that in 2017, ~1% of all residents were arrested for similar warrants (25). Of course, the proportion of defendants who are arrested for these warrants is necessarily higher than the proportion of residents, but if even 1% of defendants in our sample were arrested for failure to appear, then our interventions would have saved approximately \$140,000 from August 2016 to September 2019. Additionally, these interventions are cheap. The redesigned form has exactly the same marginal cost as the old form, and sending every summons recipient three text messages would cost New York City about \$4500 a year, making the two interventions we described here incredibly cost-effective. Due in part to these studies, all summons recipients in New York City now receive the new form and text message reminders if they provide a phone number.

These insights on the importance of insufficient awareness can likely mitigate a host of related problems in the criminal justice system. Most directly, the national failure-to-appear rate for felonies is 17 to 22% (26, 27). If insufficient awareness is partly responsible for defendants missing court for more serious offenses, there would be even greater benefits from addressing it. Even within our sample of summons recipients, we find the same treatment effects across offenses of varying severity (table S9), which suggests the possibility that insufficient awareness matters even for more serious offenses. Interventions such as the ones described here might help reduce the need for pretrial detention, as they might mitigate concerns that defendants will miss court. In fact, New York City has since expanded the use of text message reminders as a tool for reducing pretrial detention (28). Moreover, in 2016, 29% of state and federal prisoners were detained for violating some conditions of probation or parole, and it is possible that insufficient awareness might explain some of these violations (29).

More generally, our results highlight a blind spot in traditional criminal justice policies. These policies are built on an assumption that people intentionally weigh the costs and benefits

of a potential offense (30). Policies therefore often focus on deterring crime with various sanctions or punishments. Deterrence policies can only be effective when people consider the consequences of committing an offense, but this may not always happen (37). For example, bail is often used to reduce failures to appear. This system creates clear consequences for failures to appear—forfeiture of money. However, monetary bail does not appear to reduce failures to appear (32). In 2017, New York City reclassified many low-level offenses to be eligible for civil summonses (for which failures to appear do not result in warrants); yet failure-to-appear rates for these offenses did not substantially increase when the threat of warrants was removed (33). The fact that these material consequences have no effect on failures to appear suggests that many defendants are not engaged in a careful calculus of whether to skip court. Our work here goes a bit further than earlier work that has shown how this calculus might not be sensitive to punishment severity (9). Our results suggest that there may be times when such a calculus is essentially impossible because defendants are unaware of the information that would prompt (and influence) such calculation.

However, policy-makers have been slow to experiment with and adopt behavioral interventions such as these. Perhaps this is because of the criminal justice policy’s implicit assumption that failures to appear are intentional. Our laboratory experiments show that many laypeople share this assumption, and people are less supportive of these interventions when they assume that defendants intentionally skip court. It is more encouraging

that our sample of experts have different intuitions—they see how failures to appear can be unintentional, and they see a role for these interventions. However, some policies seem more closely aligned with the intuitions of our sample of nonexperts. Prior research has shown that criminal justice policies in the United States may be more aligned with popular sentiment (and perhaps more punitive) because the criminal justice system relies less on experts who are insulated from public opinion, and more on elected officials (34, 35).

Thus, widespread adoption of interventions such as these might depend on a shift in the assumptions of why failures to appear happen. Deterrence-based policies cannot be effective if people are unaware of the very information necessary for deterrence. This work suggests that making people aware of critical information may be an important addition to deterrence policies. Otherwise, these policies risk merely punishing people and not deterring actions.

Materials and methods

Form redesign details

We made four key changes to the summons form. First, the old summons form’s heading read: “Complaint/Information.” On the new form, we changed the heading to read: “Criminal Court Appearance Ticket,” to emphasize that the recipient was required to appear in court. Second, the old summons form listed the court date at the bottom. On the new form, we moved this information closer to the top of the form to make it easier to notice. Third, the old summons form required officers to write out the court’s location (again near the bottom of the form), which would have been easy to

overlook amid all of the other text on the form. On the new form, we moved this to the top, and we made it easier for officers to clearly indicate the court location. Fourth, the old summons form only noted on the back of the form that arrest warrants are issued for failures to appear. The new form included this information in bold font on the front of the form, highlighted in orange.

Message content for text message reminders

The exact wording of the text message reminders across our three treatment arms is provided in Table 1.

Laboratory experiment 1

We recruited participants from Amazon Mechanical Turk (MTurk): 232 participants completed this study. The background information on summonses included the types of offenses for which summonses are issued and the requirement to appear in court 60 to 90 days later. On the screen containing the summons form, there was a text box that reminded participants of the three pieces of information participants were searching for. When participants clicked on the corresponding information in the form, it was removed from the text box. Because reaction times are typically skewed, we analyzed the logarithm of the time it took participants to find each piece of information.

Laboratory experiment 2

We recruited 725 New York City residents from MTurk. Participants differed in many ways from the defendants in our evaluation (see table S6 for characteristics). Most notably,

Table 1. Text messages sent in the week leading up to defendants’ court dates by treatment.			
Treatment	7 days prior	3 days prior	1 day prior
Consequences	Helpful reminder: go to court Mon Jun 03 9:30AM. We’ll text to help you remember. Show up to avoid an arrest warrant. Reply STOP to end texts. www.mysummons.nyc	Remember, you have court on Mon Jun 03 at 346 Broadway Manhattan. Tickets could be dismissed or end in a fine (60 days to pay). Missing can lead to your arrest.	At court tomorrow at 9:30AM a public defender will help you through the process. Resolve your summons (ID#####) to avoid an arrest warrant.
	Helpful reminder: go to court on Mon Jun 03 9:30AM. Mark the date on your calendar and set an alarm on your phone. Reply STOP to end messages. www.mysummons.nyc	You have court on Mon Jun 03 at 346 Broadway Manhattan. What time should you leave to get there by 9:30AM? Any other arrangements to make? Write out your plan.	You have court tomorrow for summons ID#####. Did you look up directions to 346 Broadway Manhattan? Know how you’re getting there? Please arrive by 9:30AM.
Combination	Helpful reminder: go to court Mon Jun 03 9:30AM. We’ll text to help you remember. Show up to avoid an arrest warrant. Reply STOP to end texts. www.mysummons.nyc	You have court on Mon Jun 03 at 346 Broadway Manhattan. What time should you leave to get there by 9:30AM? Any other arrangements to make? Write out your plan.	Remember, you have court tomorrow at 9:30AM. Tickets could be dismissed or end in a fine (60 days to pay). Missing court for ##### can lead to your arrest.

only 4% of the MTurk respondents said they had ever received a court summons, and the sample is 60% female compared with 12% for summons recipients. However, these differences should not interact with the simple recall task in the experiment.

Participants first read a vignette in which they imagined they were involved in an altercation and received a court summons for disorderly conduct. Participants then saw their summons form, with placeholder information written into most fields. Participants were randomly assigned to see either an old or new summons form. Critically, they were informed that their court date would take place on 1 November 2017 (2 months after the incident described above) at Kings Criminal Court (346 Broadway, New York, NY 10013).

Participants then indicated the extent to which the form made them feel angry or confused (scale: strongly disagree to strongly agree). They then completed a word search as a filler task before answering questions about the forms. Next, they responded to multiple choice questions in which they were asked to recall the court date and court location listed on the form. They then indicated how likely they thought each of several outcomes would be if they missed their court date (scale: highly unlikely to highly likely): (i) The ticket would be dismissed. (ii) They would be fined. (iii) A warrant would be issued for their arrest. (iv) Nothing would happen. (v) They would get something in the mail. Finally, they responded to two multiple choice questions asking them what they were being charged with and how they could get more information.

Laboratory experiment 3

We recruited 301 U.S. residents from MTurk. Participants read five scenarios about people failing to take a required action: failing to appear for court, failing to pay an overdue bill, failing to show up for a doctor's appointment, failing to turn in paperwork for an educational program, and failing to complete a vehicle emissions test.

There are many ways in which a court appearance differs from, say, a doctor's appointment. Presumably, few people want to go to court, whereas people willingly make doctors' appointments. There are also stiff penalties for failing to appear for court, less so for other failures. And the hassles involved in attending a court appearance might be greater than the hassles involved in other actions. We attempted to control for all of these differences in our scenarios. Scenarios followed a similar template. Participants imagined that a person was required to take an action in 60 days. They were told that this person does not want to take the action, but will face a penalty for failing to do so. Participants then read that the person did not take the required action.

They then answered questions about why they thought the person failed to take action and what they think should be done to make sure that other people take the required action.

Laboratory experiment 4

We recruited 304 U.S. residents from MTurk. All participants read the same background information about summonses in New York City as in laboratory experiment 1, with additional text explaining that arrest warrants are issued for missing court (and defendants are warned about this), along with the base rate of failures to appear.

Participants were randomly assigned to one of three conditions. In the intentional condition, after reading the background information, participants responded to this question: "What is one reason why people might purposely skip their court appearance?" In the mistake condition, participants responded to this question: "What is one reason why people might accidentally miss their court appearance?" In the control condition, participants were not prompted to write anything. All participants then answered the following question: "What do you think should be done to make sure that other people show up for their court date?" Response options were: "Increase the penalty for failing to show up." "Send reminders to people about their court dates." "Make sure that court dates are easy to notice on the tickets/summonses forms issued."

Laboratory experiment 5

We recruited criminal justice experts to participate between 29 June 2020 and 10 July 2020 from several professional listservs. We received 145 complete responses (with an additional 49 partial responses). Our analyses only focus on complete responses.

After the first two parts of the experiment (described above), participants then completed several demographics questions, mostly related to their profession: what state or territory they worked in, their current role, the number of years they have been in that role, the kind of court they work in (if applicable), and how many defendants they observe failing to appear for court (if applicable). They were also asked their gender and ethnicity. Among people who completed the survey, 84 were prosecutors, 26 were defense attorneys, 7 were police officers, 7 were judges, and 21 had another profession related to criminal justice.

REFERENCES AND NOTES

1. R. Neusteter, M. O'Toole, Every three seconds: Unlocking police data on arrests (Vera Institute, 2019); www.vera.org/publication_downloads/arrest-trends-every-three-seconds-landing/arrest-trends-every-three-seconds.pdf.
2. X. Cadena, A. Schoar, "Remembering to pay? Reminders vs. financial incentives for loan repayments," Working paper 17020 (National Bureau of Economic Research (NBER) Working Paper Series, 2011); doi: [10.3386/w17020](https://doi.org/10.3386/w17020)

3. D. Karlan, M. McConnell, S. Mullainathan, J. Zinman, Getting to the top of mind: How reminders increase saving. *Manage. Sci.* **62**, 3393–3411 (2016). doi: [10.1287/mnsc.2015.2296](https://doi.org/10.1287/mnsc.2015.2296)
4. R. G. Milne, M. Horne, B. Torsney, SMS reminders in the UK national health service: An evaluation of its impact on "no-shows" at hospital out-patient clinics. *Health Care Manage. Rev.* **31**, 130–136 (2006). doi: [10.1097/00004010-200604000-00006](https://doi.org/10.1097/00004010-200604000-00006); PMID: [16648692](https://pubmed.ncbi.nlm.nih.gov/16648692/)
5. N. J. Perron *et al.*, Reduction of missed appointments at an urban primary care clinic: A randomised controlled study. *BMC Fam. Pract.* **11**, 79 (2010). doi: [10.1186/1471-2296-11-79](https://doi.org/10.1186/1471-2296-11-79); PMID: [20973950](https://pubmed.ncbi.nlm.nih.gov/20973950/)
6. P. G. Szilagyi *et al.*, Effect of patient reminder/recall interventions on immunization rates: A review. *JAMA* **284**, 1820–1827 (2000). doi: [10.1001/jama.284.14.1820](https://doi.org/10.1001/jama.284.14.1820); PMID: [11025835](https://pubmed.ncbi.nlm.nih.gov/11025835/)
7. B. L. Castleman, L. C. Page, Summer nudging: Can personalized text messages and peer mentor outreach increase college going among low-income high school graduates? *J. Law Econ. Organ.* **115**, 144–160 (2015). doi: [10.1016/j.jebo.2014.12.008](https://doi.org/10.1016/j.jebo.2014.12.008)
8. R. Paternoster, in *Deterrence, Choice, and Crime*, vol. 23, D. S. Nagin, F. T. Cullen, C. L. Jonson, Eds. (Routledge, 2019), pp. 81–106.
9. D. Nagin, Deterrence in the twenty-first century. *Crime Justice* **42**, 199–263 (2013). doi: [10.1086/670398](https://doi.org/10.1086/670398)
10. M. Patten, Q. O. Hood, C. Low-Weiner, O. Lu, E. Bond, D. Hatten, P. Chauhan, "Trends in Misdemeanor Arrests in New York 1980 to 2017" (John Jay College of Criminal Justice report, 2018); https://datacollaborativeforjustice.org/wp-content/uploads/2018/12/Executive_Summary-1.pdf.
11. S. Calonico, M. Cattaneo, R. Titiunik, Robust nonparametric confidence intervals for regression discontinuity designs. *Econometrica* **82**, 2295–2326 (2014). doi: [10.3982/ECTA11757](https://doi.org/10.3982/ECTA11757)
12. S. Calonico, M. Cattaneo, M. Farrell, R. Titiunik, Regression discontinuity designs using covariates. *Rev. Econ. Stat.* **101**, 442–451 (2019). doi: [10.1162/rest_a.00760](https://doi.org/10.1162/rest_a.00760)
13. B. H. Bornstein, A. J. Tomkins, E. M. Neeley, M. N. Herian, J. A. Hamm, Reducing courts' failure-to-appear rate by written reminders. *Psychol. Public Policy Law* **19**, 70–80 (2013). doi: [10.1037/a0026293](https://doi.org/10.1037/a0026293)
14. J. Y. Jung, B. A. Mellers, American attitudes toward nudges. *Judgm. Decis. Mak.* **11**, 62–74 (2016).
15. L. A. Reisch, C. R. Sunstein, Do Europeans like nudges? *Judgm. Decis. Mak.* **11**, 310–325 (2016). doi: [10.2139/ssrn.2739118](https://doi.org/10.2139/ssrn.2739118)
16. A. Mani, S. Mullainathan, E. Shafir, J. Zhao, Poverty impedes cognitive function. *Science* **341**, 976–980 (2013). doi: [10.1126/science.1238041](https://doi.org/10.1126/science.1238041); PMID: [23990553](https://pubmed.ncbi.nlm.nih.gov/23990553/)
17. S. Mullainathan, E. Shafir, *Scarcity: Why Having Too Little Means So Much* (Henry Holt, 2013).
18. A. K. Shah, J. Zhao, S. Mullainathan, E. Shafir, Money in the mental lives of the poor. *Soc. Cogn.* **36**, 4–19 (2018). doi: [10.1521/soco.2018.36.1.4](https://doi.org/10.1521/soco.2018.36.1.4)
19. P. Chauhan, M. Welsh, A. G. Fera, E. Balazon, "The summons report: Trends in issuance and dispositions of summonses in New York City, 2003–2013" (The Misdemeanor Justice Project, 2015); www.jjay.cuny.edu/sites/default/files/news/Summons_Report_DRAFT_4_24_2015_v8.pdf.
20. I. Kohler-Hausmann, Misdemeanor justice: Control without conviction. *Am. J. Sociol.* **119**, 351–393 (2013). doi: [10.1086/674743](https://doi.org/10.1086/674743)
21. A. Goffman, On the run: Wanted men in a Philadelphia ghetto. *Am. Sociol. Rev.* **74**, 339–357 (2009). doi: [10.1177/000312240907400301](https://doi.org/10.1177/000312240907400301)
22. S. Brayne, Surveillance and system avoidance: Criminal justice contact and institutional attachment. *Am. Sociol. Rev.* **79**, 367–391 (2014). doi: [10.1177/0003122414530398](https://doi.org/10.1177/0003122414530398)
23. N. Emanuel, H. Ho, Behavioral biases and legal compliance. *SocArXiv* (22 January 2020); <https://doi.org/10.31235/osf.io/ztrmf>.
24. T. Fain, S. Turner, M. Matsuda, "Los Angeles County Juvenile Justice Crime Prevention Act: Fiscal Year 2014–2015 Report" (Rep. no. RR-1458-LACPD, Rand Corporation, 2016); <https://doi.org/10.7249/RR1458>.
25. L. A. Slocum, B. M. Huebner, C. Greene, R. Rosenfeld, Enforcement trends in the city of St. Louis from 2007 to 2017: Exploring variability in arrests and criminal summonses over time and across communities. *J. Community Psychol.* **48**, 36–67 (2020). doi: [10.1002/jcop.22265](https://doi.org/10.1002/jcop.22265); PMID: [31705827](https://pubmed.ncbi.nlm.nih.gov/31705827/)
26. T. H. Cohen, B. A. Reaves, "Pretrial release of felony defendants in state courts" (NCJ 214994, Bureau of Justice Statistics, U.S. Department of Justice, 2007); www.bjs.gov/index.cfm?ty=pbdetail&iid=834.

27. B. A. Reaves, "Felony defendants in large urban counties, 2009 - Statistical tables" (NCJ 243777, Bureau of Justice Statistics, U.S. Department of Justice, 2013); www.bjs.gov/index.cfm?ty=pbdetail&iid=4845.
28. R. Ferri, The benefits of live court date reminder phone calls during pretrial case processing. *J. Exp. Criminol.* 10.1007/s11292-020-09423-0 (2020). doi: [10.1007/s11292-020-09423-0](https://doi.org/10.1007/s11292-020-09423-0)
29. E. A. Carson, "Prisoners in 2016" (NCJ 251149, Bureau of Justice Statistics, U.S. Department of Justice, 2018); www.bjs.gov/index.cfm?ty=pbdetail&iid=6187.
30. G. S. Becker, Crime and punishment: An economic approach. *J. Polit. Econ.* **76**, 169–217 (1968). doi: [10.1086/259394](https://doi.org/10.1086/259394)
31. A. K. Shah, J. Ludwig, Option Awareness: The psychology of what we consider. *Am. Econ. Rev.* **106**, 425–429 (2016). doi: [10.1257/aer.p20161098](https://doi.org/10.1257/aer.p20161098)
32. A. Ouss, M. T. Stevenson, Bail, jail, and pretrial misconduct: The influence of prosecutors. SSRN 3335138 [Preprint]. 20 June 2020; doi: [10.2139/ssrn.3335138](https://doi.org/10.2139/ssrn.3335138)
33. C. Cuevas, E. Grimsley, P. Chauhan, K. Mulligan, Criminal and Civil Court Appearance: Predictors of Timely Response to Summonses for Lower-Level Offenses in New York City. *Criminol., Criminal Just., Law & Soc.* **20**, 1–24 (2019).
34. S. C. Gordon, G. A. Huber, The effect of electoral competitiveness on incumbent behavior. *Quarterly J. Pol. Sci.* **2**, 107–138 (2007). doi: [10.1561/100.00006035](https://doi.org/10.1561/100.00006035)

35. F. E. Zimring, Protect individual punishment decisions from mandatory penalties. *Criminol. Public Policy* **6**, 881–886 (2007). doi: [10.1111/j.1745-9133.2007.00468.x](https://doi.org/10.1111/j.1745-9133.2007.00468.x)

ACKNOWLEDGMENTS

We greatly appreciate the support of the New York City Mayor's Office of Criminal Justice, in particular E. Glazer, A. Crohn, A. Meizlish, and A. LaScala-Gruenewald; the New York City Police Department, in particular Deputy Commissioner S. Herman, Detective K. Rice, and Lieutenant D. O'Hanlon; and the New York State Unified Court System Office of Court Administration, in particular J. Barry, J. Hill, K. Kane, C. Cadoret, and Z. Bedell. Any data provided herein does not constitute an official record of the New York State Unified Court System, which does not represent or warrant the accuracy thereof. The opinions, findings, and conclusions expressed in this publication are those of the authors and not those of the New York State Unified Court System, which assumes no liability for its contents or use thereof. We thank J. Ludwig and S. Mullainathan for helpful comments. We thank C. Avellan, M. Bluestine, H. Furstenberg-Beckman, J. Leifer, K. Lorenzana, S. McAuliffe, R. Ander, K. B. Falco, B. Z. Diop, E. Fletcher, C. Hanlock, J. Lefkowitz, D. Mungia Gomez, E. Packis, B. Pinder, K. Pinder, A. Yates-Berg, and Z. Honoroff for invaluable project assistance. We especially thank B. Cooke and J. Hayes for their work on the design of the text message intervention and experiment and for their outstanding management of the project

and partners. We also thank C. Burke for graphic design assistance. This research was approved by the University of Chicago's Institutional Review Board (protocols IRB15-0206 and IRB17-0928). **Funding:** This work was supported with grants from the Laura and John Arnold Foundation, the John D. and Catherine T. MacArthur Foundation, and the Abdul Latif Jameel Poverty Action Lab (J-PAL). **Author contributions:** All authors contributed to the research design, to reviewing the findings, and to writing the manuscript. **Competing interests:** The authors have no competing interests. **Data and materials and availability:** Preregistration details, all analysis code, data from the laboratory experiments, and anonymized data allowing researchers to reproduce results from the field experiments are available on OSF (<https://osf.io/mb6x7/>).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6517/eabb6591/suppl/DC1
 Supplementary Materials and Methods
 Supplementary Text
 Figs. S1 to S9
 Tables S1 to S13
 References (36–38)
 MDAR Reproducibility Checklist
 10 March 2020; accepted 24 September 2020
 Published online 8 October 2020
 10.1126/science.abb6591

RESEARCH ARTICLE

METABOLIC GENOMICS

Resource conservation manifests in the genetic code

Liat Shenhav^{1,2*} and David Zeevi^{1*†}

Nutrient limitation drives competition for resources across organisms. However, much is unknown about how selective pressures resulting from nutrient limitation shape microbial coding sequences. Here, we study this “resource-driven selection” by using metagenomic and single-cell data of marine microbes, alongside environmental measurements. We show that a significant portion of the selection exerted on microbes is explained by the environment and is associated with nitrogen availability. Notably, this resource conservation optimization is encoded in the structure of the standard genetic code, providing robustness against mutations that increase carbon and nitrogen incorporation into protein sequences. This robustness generalizes to codon choices from multiple taxa across all domains of life, including the human genome.

Nitrogen and carbon are major limiting factors in many ecosystems, with recent studies linking their availability to core genomic properties (1, 2). In low-nitrogen environments, there is a strong A+T bias in nucleotide sequences, smaller genome sizes, and a lower incorporation of nitrogen-rich side chains into proteins (3–5). Opposite trends have been shown for carbon limitation (2), and indeed, nitrogen and carbon concentrations are typically inversely correlated (6). Thus, recent studies propose a purifying selective pressure associated with resource conservation (2, 3, 5), which we term “resource-driven selection.”

Resource-driven selection postulates that mutations resulting in excess incorporation of nutrients such as nitrogen and carbon are disfavored. However, not all mutations have the same effect on protein sequences, because of constraints imposed by the pattern of codon assignments in the standard genetic code (hereinafter, the structure of the genetic code). The genetic code, common to virtually all of life on earth, can mitigate the effects of mistranslation errors and point mutations (7), specifically those leading to radical changes in amino acids. This error minimization is prominent among theories regarding the origin of the genetic code (8–11), which propose that the code evolved through selection to minimize potential adverse effects of mutations on protein structure and function (12–14). To quantify code optimality, some theories provide structurally informed amino acid metrics on the basis of hydropathy and stereochemistry [e.g., the polar requirement (PR) scale (11) and hydrophathy index (15)]. To our knowledge, an opti-

mization of nutrient conservation in the genetic code has not been studied thus far.

Results

Widespread purifying selection in the marine environment

To comprehensively characterize how coding sequences of marine microbes are affected by resource availability, we first downloaded 746 samples from the Tara Oceans consortium ($n = 136$) (16), bioGEOTrACES ($n = 480$) (17), and the Hawaii Ocean time series (HOT; $n = 68$) and Bermuda Atlantic time series (BATS; $n = 62$) (17) (fig. S1A) (18). We then devised a computational pipeline that calculates selection metrics from these marine metagenomic samples (fig. S1B). We aligned reads to the Ocean Microbiome Reference Gene Catalog (OM-RGC) (16), a database of genes from marine envi-

ronments that is accompanied by functional information. We searched for single-nucleotide polymorphisms (SNPs) in genes that had sufficient high-quality coverage (fig. S1B) (18). Overall, we found 71,921,864 high-confidence SNPs (18), in a total of 1,590,843 genes.

To quantify purifying selection on different gene functions, we annotated genes from the OM-RGC database using either KEGG orthology groups (KOs) (19) or eggNOG orthologous groups (OGs) (18, 20). Using called SNPs, we calculated for each orthologous group and in each sample the ratio of nonsynonymous polymorphisms to synonymous polymorphisms (pN/pS) (18). Across all samples, we found pN/pS ratios to be close to zero with an average of 0.074 in eggNOG OGs [confidence interval (CI), 0.072 to 0.075] (fig. S1C) and 0.079 in KEGG KOs (CI, 0.077 to 0.080) (fig. S1D), indicating purifying selection across the marine environment. To corroborate the validity of calculating selection metrics from metagenomic samples, we compared nonsynonymous mutations leading to “conservative” amino acid substitutions with those leading to “radical” substitutions and found conservative mutations to be significantly more common (permutation test, $P < 0.0001$) (21).

Resource-driven selection apparent across marine microbial genes

On the basis of recent studies (2–5), we hypothesized that nutrient availability is a central driver of this purifying selection. We thus considered environmental measurements taken alongside each sample, including depth, water temperature, salinity, and concentrations of nitrate, nitrite, oxygen, phosphate, and silicate (fig. S2, A to H) (18).

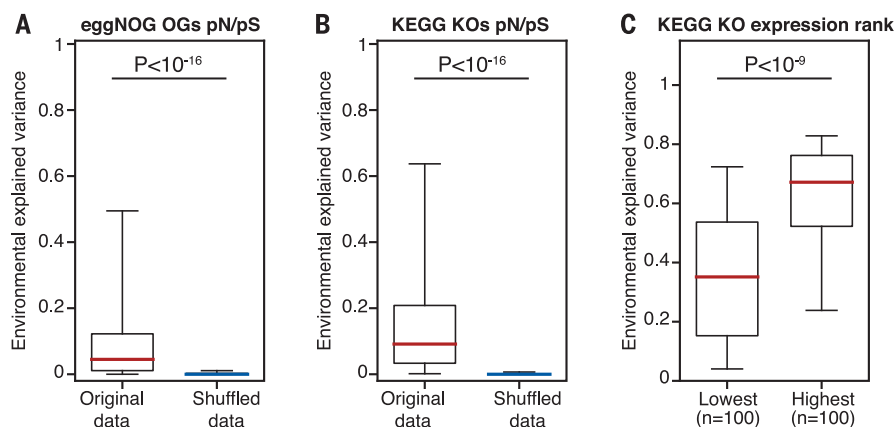


Fig. 1. Analysis of pN/pS ratios reveals resource-driven selection. (A) Variance of eggNOG OG pN/pS explained by the environment in the LMM (red) (18) compared with the same data with shuffled labels (blue). The box plot lines show median values, the boxes show interquartile ranges, and whiskers show 5th and 95th percentiles. (B) Similar data presentation as in (A), for KEGG KO pN/pS. P values were determined with the Wilcoxon signed-rank test. (C) Box plot of variance in pN/pS explained by the environment in the LMM in the 100 lowest (left) and 100 highest (right) expressed KEGG KOs. P values were determined with the Mann-Whitney U test.

¹Center for Studies in Physics and Biology, Rockefeller University, New York, NY, USA. ²Department of Computer Science, University of California Los Angeles, Los Angeles, CA, USA.

*These authors contributed equally to this work.

†Corresponding author. Email: dzeevi@rockefeller.edu

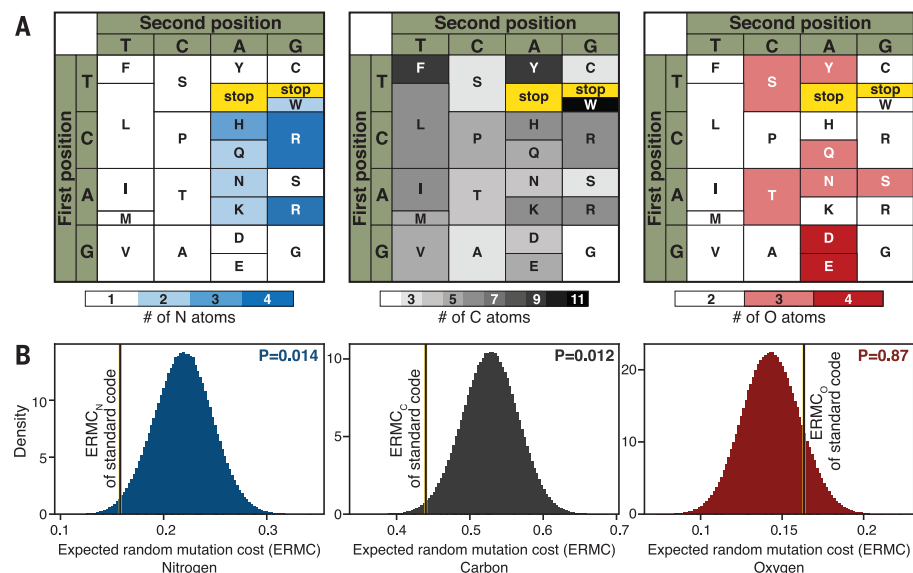


Fig. 2. Resource conservation is facilitated by the genetic code. (A) Nitrogen (left), carbon (center), and oxygen (right) content of different amino acids, depicted along with their positions in the standard genetic code. (B) Histograms of the ERMC in 1 million random permutations of the genetic code for nitrogen (blue), carbon (black), and oxygen (red). The black and yellow bars mark the ERMCs of the standard genetic code for each of the elements. *P* values were determined with a permutation test.

These measurements presented consistent correlation patterns with the pN/pS ratios of many KEGG and eggNOG orthologs (fig. S3). However, as they are also correlated with each other (fig. S2I), we cannot accurately estimate their individual effects. We therefore used a linear mixed model (LMM) with variance components (18) to estimate the fraction of variance in pN/pS ratios (dependent variable) that is explained by the environment (random effect), while controlling for the correlation structure between the environmental parameters. We term the fraction of variance in pN/pS ratios explained by resource availability the environmental explained variance (EEV) (18). Across both KEGG and eggNOG orthologs, a significant fraction of the variance in pN/pS can be attributed to the environment (Mann-Whitney *U* test, $P < 10^{-16}$) (Fig. 1, A and B, and fig. S4, A and B), with nitrate being more strongly correlated with pN/pS ratios than any other environmental parameter (Kolmogorov-Smirnov test, $P < 10^{-30}$ for all comparisons) (fig. S5). Examining typical DNA mutations and amino acid substitutions in nitrate-rich versus nitrate-poor environments, we found that environmental nitrate is associated with specific changes to both DNA and protein sequences, favoring lower nitrogen incorporation into protein sequences when nitrate is scarce (21).

This association between environmental measurements and the magnitude of purifying selection is significant even after controlling for potential confounders such as time or effective population size (Mann-Whitney *U* test, $P < 10^{-16}$)

(fig. S4, C and D) (21), as well as in specific environmental niches (Mann-Whitney *U* test, $P < 10^{-20}$) (fig. S6) (21). Additionally, these results were replicated using benchmarking data of assembled genomes from uncultivated single cells from three dominant lineages of the surface ocean (SAR-11, SAR-86, and *Prochlorococcus*) (fig. S4, E to G) (21, 22). These validations demonstrated that the association between selective pressure and environmental conditions is robust to both data type and selection metric and is not confounded by population properties and clade-specific metabolism.

Environmental association is stronger in resource-consuming genes

With nitrate being the environmental factor most strongly associated with pN/pS, we and others hypothesize that mutations that increase the nitrogen requirements of cells are selected against, especially in nitrogen-limited conditions (3, 5, 23). This implies stronger purifying selection in highly expressed genes (3, 4), where one DNA mutation could translate to thousands of proteins, each consuming more resources (illustrated in fig. S7). We thus used an expression dataset for marine microbial genes (24) to rank KEGG KOs by their mean expression (18). The 100 most highly expressed KEGG KOs had a significantly higher EEV than the 100 least-expressed ones (Mann-Whitney *U* test, $P < 10^{-9}$) (Fig. 1C and fig. S8). We replicated these results using single-cell data pertaining to specific bacterioplankton lineages (Mann-Whitney *U* test, $P < 10^{-7}$). Additionally, we found that genes encod-

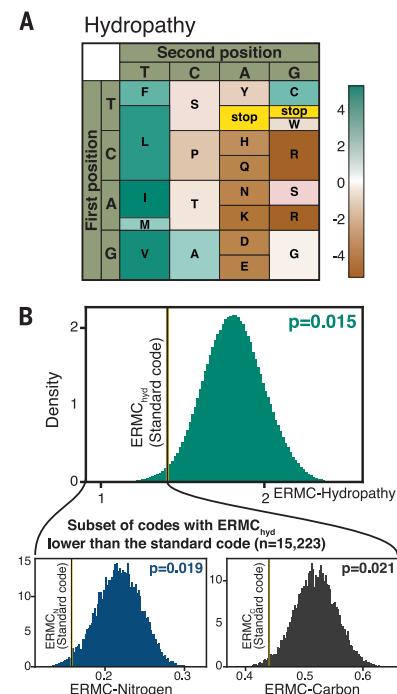


Fig. 3. Optimization for carbon and nitrogen is not confounded by hydrophathy. (A) Hydrophathy of different amino acids, depicted along their positions in the standard genetic code. (B, top) Histograms of ERMCs in 1 million random permutations of the genetic code for hydrophathy. (Bottom) Histograms of ERMCs for nitrogen (left) and carbon (right) for the subset of hypothetical genetic codes with ERMCs lower than the standard genetic code (n=15,223). *P* values were determined with a permutation test.

ing extracellular proteins, i.e., for resources excreted from the cell that cannot be recycled, had significantly higher EEV than other gene groups (Mann-Whitney *U* test, $P < 0.05$) (fig. S9) (18). This higher EEV for resource-consuming genes further strengthens our results regarding the breadth of resource-driven selection.

Resource conservation as an optimization mechanism in the genetic code

We observed selection against DNA mutations which result in excess incorporation of nutrients, such as nitrogen and carbon, into proteins. However, mutations are constrained by the structure of the genetic code, which minimizes the impact of point mutations on protein structure and function (12–14). We hypothesized that the genetic code also minimizes the impact of point mutations on nutrient incorporation into proteins. Specifically, the genetic code acts as a buffer between DNA, where mutations occur, and proteins, where resource-driven selection is exerted.

We thus defined, for each element *e* (e.g., carbon, nitrogen), a function quantifying the cost of a single mutation as the added number of amino acid atoms resulting from it.

For example, a missense mutation from codon CCA to CGA results in an amino acid substitution from proline to arginine, with an increase of one carbon and three nitrogen atoms, setting the nitrogen cost of such a mutation to 3 and the carbon cost to 1 (Fig. 2A).

We calculated, for nitrogen, carbon, and oxygen, the expected random mutation cost (ERMC) for the standard genetic code. This calculation considered the abundance of codons, calculated from all marine samples; the transition probability between codons, estimated using the abundance of all single-nucleotide mutations (e.g., for the mutation changing GCA to CCA, we use the abundance of G-to-C transversions); and the cost function, i.e., the number of atoms of each element added after mutation (18). For the standard genetic code, and also for codon abundances and mutation rates calculated for marine microbes, we report ERMC values of 0.44, 0.16, and 0.16 for carbon, nitrogen, and oxygen, respectively, corresponding to an average increase of this number of atoms per random mutation (Fig. 2B).

To check if the standard genetic code, along with codon abundances and mutation rates, is indeed robust to resource-consuming mutations, we compared it with other hypothetical codes. We simulated alternative genetic codes by randomizing the first and second position in all codons 1 million times, to create a null distribution of ERMC (18). We found that the standard genetic code, common to most life forms, is parsimonious in terms of carbon and nitrogen utilization, given a random mutation. This is marked by a significantly low ERMC for nitrogen ($\text{ERMC}_N P = 0.014$) (Fig. 2B) and carbon ($\text{ERMC}_C P = 0.012$), but not oxygen ($\text{ERMC}_O P = 0.87$).

We compared the extent of robustness against addition of carbon or nitrogen to protein sequences with the robustness against amino acid changes that may affect protein structure and function, as extensively reported (12–14). We calculated ERMC for changes in hydrophathy (15) and polar requirement (PR) (11), both of which are structurally informed amino acid properties used to determine code error minimization (18). We found that these optimization mechanisms are of a similar magnitude of significance as those for nitrogen and carbon conservation ($\text{ERMC}_{PR} P = 0.014$; $\text{ERMC}_{hyd} P = 0.015$) (fig. S10, C and D) (18).

We then devised a hierarchical model to examine the subset of genetic codes (out of 1 million hypothetical codes tested) that have a lower ERMC than the standard code for PR or hydrophathy, and we tested whether this subset is also optimized for nitrogen or carbon (18). If nutrient optimization is separate from structural optimization, we would expect the standard code to be optimized for carbon and nitrogen, even in comparisons with

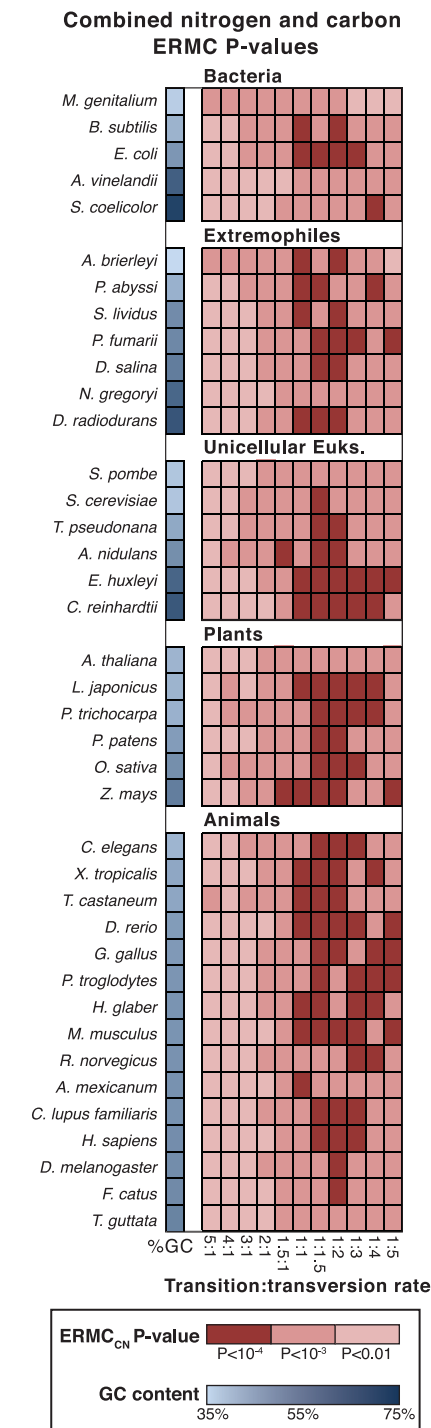


Fig. 4. The genetic code is optimized for resource conservation across organisms. Heatmap of ERMC_{CN} P values across 39 organisms and 11 transition:transversion rates. Organisms in each group are ordered by the GC content of their coding sequences.

this subset, with significantly lower ERMC_N and ERMC_C values. Of the 15,223 hypothetical codes that have a hydrophathy ERMC (ERMC_{hyd}) lower than the standard code, only 270 have a lower ERMC_N ($P = 0.019$) (Fig. 3 and fig. S10E)

and only 249 have a lower ERMC_C ($P = 0.021$) (Fig. 3 and fig. S10E). Similarly, of the 13,729 hypothetical codes that have an ERMC_{PR} lower than the standard code, only 83 have a lower ERMC_N ($P = 0.006$) (fig. S10F), and only 442 have a lower ERMC_C ($P = 0.037$) (fig. S10F). This is in contrast with the observed overlap between hydrophathy and PR: out of the 15,223 hypothetical codes that have an ERMC_{hyd} lower than the standard code, 6736 have a lower ERMC_{PR} ($P = 0.44$) (fig. S10E). These results indicate that the detected carbon and nitrogen optimization is not confounded by previously reported optimization properties such as hydrophathy and PR.

Remarkably, only 128 out of 1 million randomized genetic codes were better than the standard code in conservation of nitrogen and carbon together ($\text{ERMC}_{CN} P = 1.3 \times 10^{-4}$) (fig. S10, G and H). This number is significantly smaller than the number of hypothetical codes expected to have both a lower ERMC_C and ERMC_N (chi-square test of independence, $P = 0.0013$) (table S1). This is possibly driven by a small overlap between the positions of high-nitrogen and high-carbon amino acids. This property of the standard code potentially enables concurrent optimization for both carbon and nitrogen. These results highlight a new optimization principle of the genetic code that is of similar magnitude—and independent of—previously proposed principles.

The genetic code facilitates resource conservation across kingdoms

To show that the resource robustness of the genetic code was not limited to our dataset, we calculated the ERMC of 187 strains of marine microbes in the genera *Prochlorococcus* and *Synechococcus*. We computed codon abundances and mutation rates using published protein-coding sequences (1) and the accepted transition:transversion rate of 2:1 (18, 25). We identified significant conservation of carbon, nitrogen, and both elements combined (ERMC_C mean, $P = 0.013$ and $P = 0.020$; ERMC_N mean, $P = 0.049$ and $P = 0.032$; $\text{ERMC}_{CN} P = 0.0004$, $P = 0.0007$ for *Prochlorococcus* and *Synechococcus*, respectively) (fig. S11A).

To explore whether this nutrient conservation optimization in the genetic code extends across organisms, we performed a similar calculation using codon abundances from 39 organisms across all domains of life, including all human protein-coding sequences and a range of transition:transversion rates (18). Similarly to marine microbes, we found that the genetic code is optimized in terms of resource utilization for all tested organisms, and it is marked by a significantly lower ERMC for nitrogen and carbon combined, across all transition:transversion rates ($P < 0.01$) (Fig. 4). Moreover, we found significant optimization even in the theoretical case where all codon

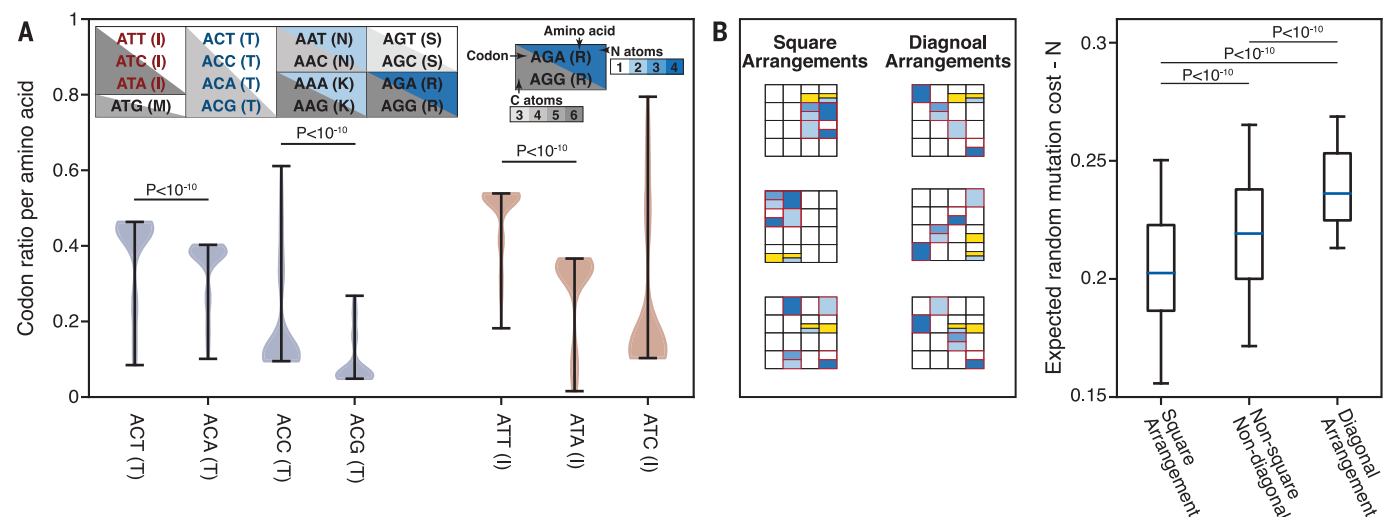


Fig. 5. Structural properties and codon usage bias underlying optimality in the genetic code. (A) Violin plot of codon usage among 187 species of *Prochlorococcus* and *Synechococcus*, showing significant preference of threonine codons ACT and ACC compared with ACA and ACG and of isoleucine codon ATT compared with ACA. P values were determined with the Wilcoxon signed-rank

test. (B) ERM_{CN} values for square arrangements (left) and diagonal arrangements (right) (18), compared with all other arrangements (center) out of 10,000 hypothetical arrangements. The box plot lines show median values, the boxes show interquartile ranges, and whiskers show 5th and 95th percentiles. P values were determined with the Mann-Whitney U test.

abundances are the same ($P < 0.01$) (fig. S11B). The codon abundances of a great majority of organisms also demonstrate significantly lower ERM_{CN} values for nitrogen (fig. S11C) and carbon (fig. S11D), for a wide range of transition:transversion rates. These results indicate that resource optimization in the genetic code transcends taxonomy, codon choices, and mutation rates.

Resource conservation may bias codon usage

We examined all amino acids encoded by codons with adenine in the first position, focusing on codon usage of the amino acid threonine. We note that a C-to-G transversion in the second position for codons ACT and ACC yields serine (AGT and AGC, respectively), but the same mutation for codons ACA and ACG yields arginine (AGA and AGG) (Fig. 5A, inset). Arginine has higher carbon and nitrogen contents than serine. We thus hypothesized that for a cell to conserve nutrients in case of a random mutation, codons ACT and ACC should have a higher abundance than codons ACA and ACG, respectively, given a known genomic GC bias. We examined codon usage in 187 *Prochlorococcus* and *Synechococcus* strains and found significantly higher use of ACT than of ACA (Wilcoxon signed-rank test, $P < 10^{-20}$) (Fig. 5A) and significantly higher use of ACC than of ACG (Wilcoxon signed-rank test, $P < 10^{-20}$) (Fig. 5A). Similarly, the isoleucine codon ATT had a higher abundance than ATA (Wilcoxon signed-rank test, $P < 10^{-20}$) (Fig. 5A). These results point to resource conservation as a central driving force in guiding codon usage and thereby affecting not only protein sequence but also cellular translation efficiency.

Structural principles drive optimization in the genetic code

Codons of the nitrogen-rich amino acids histidine, glutamine, asparagine, lysine, and arginine span only two nucleotides in their first position and two in their second position. We define this organization to be a “square” arrangement and hypothesize that it amplifies nitrogen conservation (Fig. 5B) (18). Specifically, in the square arrangement, some codons require at least two mutations to increase the number of nitrogen atoms (e.g., those coding for alanine and valine). This is in contrast to other hypothetical arrangements, including a “diagonal” one in which nitrogen-rich amino acid codons span all possible nucleotides in the first and second positions (Fig. 5B) (18). On the basis of these results, it seems that the diagonal arrangement would be nutrient-wasteful, as a single mutation in the first or second codon position could increase the nitrogen content of a protein. We generated 10,000 hypothetical codes, with 220 arrangements embodying a square structure and 127 embodying a diagonal one. Compared with all other possible arrangements, the square arrangements present a significantly lower ERM_{CN}, whereas diagonal arrangements exhibit a significantly higher ERM_{CN} (Mann-Whitney U test, $P < 10^{-10}$ for both) (Fig. 5B). This indicates that resource optimization in the standard code is driven by structural principles, perhaps underlying the optimization observed across kingdoms.

Discussion

Here, we characterized and quantified the selective forces exerted by nutrient availability

on protein-coding genes in marine environments. We provide a data-driven, population-level perspective and show that resource-driven selection is a ubiquitous force. We further show that a significant portion of DNA mutations may not result in increased nutrient incorporation into protein sequences, owing to the pattern of codon assignments in the genetic code.

In light of these results, we hypothesize that resource-driven selection is equally exerted on all parts of the protein-coding gene. This sets it apart from selection to maintain the structural integrity of a protein or the function of its active site, both of which occur predominantly in structurally important regions (26). Thus, accounting for resource-driven selection may improve the identification of alternative translation start sites, alternatively spliced introns, or readthrough stop codons, as intermittently translated regions of the protein may be under weaker resource-driven selection than are constitutively translated ones (27).

Our results showing that the genetic code optimizes nutrient conservation are in line with theories sporting an early fixation of an optimal genetic code, suggesting selection for error minimization (10, 28). Nevertheless, the genetic code is also near-immutable, as evident in heterotrophic eukaryotes, which as net nitrogen producers still harbor a nitrogen-conservative code. This implies that the genetic code can be viewed as a buffer between the evolutionary forces of mutation and selection, the former occurring in DNA sequences and the latter predominantly in proteins. In the case of nutrient conservation, many DNA mutations do not result in the incorporation

of additional nutrients into proteins and are thus not selected against, which may allow more “freedom” for fitness gradients to explore the mutation space.

REFERENCES AND NOTES

1. P. M. Berube, A. Rasmussen, R. Braakman, R. Stepanauskas, S. W. Chisholm, *eLife* **8**, e41043 (2019).
2. D. R. Mende *et al.*, *Nat. Microbiol.* **2**, 1367–1373 (2017).
3. J. J. Grzymalski, A. M. Dussa, *ISME J.* **6**, 71–80 (2012).
4. H. Akashi, T. Gojobori, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 3695–3700 (2002).
5. F. L. Hellweger, Y. Huang, H. Luo, *ISME J.* **12**, 1180–1187 (2018).
6. E. R. M. Druffel, S. Griffin, A. I. Coppola, B. D. Walker, *Geophys. Res. Lett.* **43**, 5279–5286 (2016).
7. R. T. Hinegardner, J. Engelberg, *Science* **142**, 1083–1085 (1963).
8. G. Gamow, *Nature* **173**, 318 (1954).
9. J. T. Wong, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 1909–1912 (1975).
10. S. J. Freeland, L. D. Hurst, *J. Mol. Evol.* **47**, 238–248 (1998).
11. C. R. Woese, D. H. Dugre, W. C. Saxinger, S. A. Dugre, *Proc. Natl. Acad. Sci. U.S.A.* **55**, 966–974 (1966).
12. S. E. Massey, *J. Mol. Evol.* **67**, 510–516 (2008).
13. A. S. Novozhilov, E. V. Koonin, *Biol. Direct* **4**, 44 (2009).
14. N. Torabi, H. Goodarzi, H. Shateri Najafabadi, *J. Theor. Biol.* **244**, 737–744 (2007).
15. D. Haig, L. D. Hurst, *J. Mol. Evol.* **33**, 412–417 (1991).
16. S. Sunagawa *et al.*, *Science* **348**, 1261359 (2015).
17. S. J. Biller *et al.*, *Sci. Data* **5**, 180176 (2018).
18. Materials and methods are available as supplementary materials.
19. M. Kanehisa, S. Goto, *Nucleic Acids Res.* **28**, 27–30 (2000).
20. J. Huerta-Cepas *et al.*, *Mol. Biol. Evol.* **34**, 2115–2122 (2017).
21. Supplementary text is available as supplementary materials.
22. M. G. Pachadaki *et al.*, *Cell* **179**, 1623–1635.e11 (2019).
23. A. Dufresne, L. Garczarek, F. Partensky, *Genome Biol.* **6**, R14 (2005).
24. B. C. Kolody *et al.*, *ISME J.* **13**, 2817–2833 (2019).
25. E. Urbach, D. J. Scanlan, D. L. Distel, J. B. Waterbury, S. W. Chisholm, *J. Mol. Evol.* **46**, 188–201 (1998).
26. J. Liu, Y. Zhang, X. Lei, Z. Zhang, *Genome Biol.* **9**, R69 (2008).
27. R. W. Read *et al.*, *ISME J.* **11**, 2267–2278 (2017).
28. S. J. Freeland, R. D. Knight, L. F. Landweber, L. D. Hurst, *Mol. Biol. Evol.* **17**, 511–518 (2000).
29. L. Shenhav, D. Zeevi, Resource conservation manifests in the genetic code, Zenodo (2020); doi: 5281/zenodo.4001183.

ACKNOWLEDGMENTS

We thank T. Korem and N. D. Rubinstein for discussions, suggestions, and help in analyses. We thank J. P. Eckmann, O. X. Cordero, I. Mizrahi, and N. Antonovsky for their suggestions. **Funding:** D.Z. was funded by the James S. McDonnell Foundation. **Author contributions:** L.S. and D.Z. conceived the project, performed analyses, and wrote the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the manuscript, supplementary materials, and referenced datasets. Code is available at Zenodo (29).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6517/683/suppl/DC1
Materials and Methods
Supplementary text
Figs. S1 to S16
Tables S1 to S4
References (30–56)
MDAR Reproducibility Checklist

23 October 2019; resubmitted 4 May 2020
Accepted 11 September 2020
10.1126/science.aaz9642

PALEONTOLOGY

Enigmatic amphibians in mid-Cretaceous amber were chameleon-like ballistic feeders

Juan D. Daza^{1*}, Edward L. Stanley², Arnau Bolet^{3,4}, Aaron M. Bauer⁵, J. Salvador Arias⁶, Andrej Čerňanský⁷, Joseph J. Bevirt⁸, Philipp Wagner⁹, Susan E. Evans¹⁰

Albanerpetontids are tiny, enigmatic fossil amphibians with a distinctive suite of characteristics, including scales and specialized jaw and neck joints. Here we describe a new genus and species of albanerpetontid, represented by fully articulated and three-dimensional specimens preserved in amber. These specimens preserve skeletal and soft tissues, including an elongated median hyoid element, the tip of which remains embedded in a distal tongue pad. This arrangement is very similar to the long, rapidly projecting tongue of chameleons. Our results thus suggest that albanerpetontids were sit-and-wait ballistic tongue feeders, extending the record of this specialized feeding mode by around 100 million years.

The extinct amphibian clade Albanerpetontidae is currently represented by five genera spanning a period of more than 165 million years, from the Middle Jurassic (1) to the Early Pleistocene (2), and a geographical range from North America (3, 4) through Europe (5) and central Asia (6) to Japan (7). To date, the only records from a southern (Gondwanan) continental mass, albeit marginal, are from Morocco (8, 9). Most of albanerpetontid records comprise jaws and sculptured frontal bones (5), although Cretaceous localities [in Spain and Italy (1, 10, 11) and Japan (7)] have yielded more substantive material, some with associated soft tissue showing the presence of dermal scales (10). Originally classified as salamanders (11, 12), albanerpetontids are now considered a distinct lineage (10, 13). Nonetheless, many questions remain as to the anatomy, lifestyle, and relationships of these unusual amphibians.

Here we describe a new genus and species of albanerpetontid from the amber deposits of Myanmar. The material includes a complete three-dimensional adult skull, a tiny juvenile skeleton originally identified as a putative stem-chameleon (14), and a partial adult postcranium. The exquisite preservation of both skeletal remains and soft tissues reveals important clues

about the morphological and ecological character of these enigmatic amphibians.

Systematics

Amphibia Linnaeus 1758. Albanerpetontidae Fox & Naylor 1982. *Yaksha perettii* gen. et sp. nov. **LSID** urn:lsid:zoobank.org:pub:0C8EC7C5-66D4-4144-917D-5BFA3704EFA4. **Etymology:** The generic name is derived from Yaksha, a type of mythical spirit in Eastern belief systems, guardian of natural treasures hidden in the earth or tree roots. The specific epithet recognizes Adolf Peretti, director of the Peretti Museum Foundation and GemResearch Swisslab (GRS), who discovered the fossil and has conducted fieldwork and humanitarian projects in Myanmar for the past 10 years. **Holotype:** Peretti Museum Foundation, Switzerland, GRS-Ref-060829 (15), a complete articulated skull (Figs. 1, A and G to N; 2, A and B; and 3). **Paratype:** JZC Bu154 (16) [Fig. 1, B to F; figs. S2 and S3; and figure 2K in Daza *et al.* (14)], James Zigras Collection, juvenile specimen, housed at the American Museum of Natural History, New York, USA. **Referred material:** Peretti Museum Foundation, Switzerland, GRS-Ref-27746 (17), partial postcranial skeleton (fig. S5). **Locality and horizon:** GRS-Ref-060829 (holotype) and GRS-Ref-27746 (referred material) were ethically sourced from the Hukaung Valley, near Tanaing Township, Myitkyina District, Kachin Province, Myanmar, and legally exported (materials and methods section S1.1 in the supplementary materials). The juvenile paratype, JZC Bu154, is recorded as being from ~100 km west of the Myitkyina District. Amber from these mines has been dated as early Cenomanian, ~99 million years ago (Ma), using U-Pb isotopes (18). **Diagnosis for genus and species:** A genus and species of albanerpetontid distinguished by the following combination of character states: paired robust premaxillae with a dorsal boss, wide lateral lingual buttress, and elongate vertical suprapalatal pits (Fig. 3, A to C); posteriorly

bifurcate parietals bounding cranial fenestrations anteriorly and medially (Fig. 3, D and E); triangular frontal with long broad-based internasal process, frontal anteroposterior length equal to maximum anteroposterior length of parietal (Fig. 3, F to H), prefrontal facets extending posterior to midlength of frontal, weakly developed midventral crest, and ventrolateral crests that meet in ventral midline; medium-length parietal postorbital processes sculptured in their proximal half; separate prefrontal and lacrimal bones; nasal excluded from narial margin; trifurcate unpaired vomer; dentition showing size heterodonty anteriorly, resulting in sinuous occlusal surface; and small body size (supplementary text section S2.1, “Differential diagnosis”).

Description

The holotype (GRS-Ref-060829) is the first fully articulated three-dimensional skull of an albanerpetontid (Fig. 1, A and G to N, and supplementary text S2.2). It is 12.18 mm in overall length (snout tip to occiput), giving an estimated snout-to-pelvis length (SPL) of 52 mm [based on the proportions of *Celtedens ibericus* (10)]. The newly discovered Myanmar skull reveals the presence of epipterygoids [misinterpreted in the Japanese *Shirerpeton* (7)] that form an integral part of the jaw suspension; a complete braincase with ossification of the pila antoticae; a palate with a large, open pyriform fossa undivided by a median parasphenoid rostrum; a trifurcated unpaired vomer, with paired palatines and pterygoids, all lacking teeth; and a long median hyoid “entoglossal” process (not homologous with that of lizards). There are no ossified ceratobranchial elements. GRS-Ref-060829 also preserves remnants of the original soft tissues, notably the anterior tongue pad, into which the tip of the entoglossal process is embedded, and parts of the eyelids, palatal fascia, and jaw musculature.

The juvenile specimen, JZC Bu154 (Fig. 1, B to F, and supplementary text S2.3), establishes the presence of a four-digit manus, uncertain in other articulated specimens (1, 10), and curved ungual phalanges covered by claw sheaths. GRS-Ref-27746 (fig. S5 and supplementary text S2.4) demonstrates that the tripartite pelvis (pubis, ischium, and vertical ilium) was supported by two sacral ribs (unlike the single sacral rib typically found in lissamphibians). Moreover, the vertical iliac blade, like that of chameleons, suggests a deep pelvis that may have allowed the legs to be angled ventrally for climbing.

Phylogenetic position

Most albanerpetontid species are represented by isolated bones, and data matrices for the group rely mainly on frontal and jaw character states. We ran an analysis using the most comprehensive recent data matrix (7, 19). In their possession of dorsal cranial fenestrae, *Yaksha*

¹Department of Biological Sciences, Sam Houston State University, Huntsville, TX, USA. ²Department of Herpetology, Florida Museum of Natural History, Gainesville, FL, USA.

³Institut Català de Paleontologia Miquel Crusafont, Universitat Autònoma de Barcelona, Cerdanyola del Vallès, Spain.

⁴School of Earth Sciences, University of Bristol, Bristol, UK.

⁵Department of Biology and Center for Biodiversity and Ecosystem Stewardship, Villanova University, Villanova, PA, USA. ⁶Unidad Ejecutora Lillo, CONICET - FML, San Miguel de Tucumán, Argentina. ⁷Department of Ecology, Laboratory of Evolutionary Biology, Faculty of Natural Sciences, Comenius University in Bratislava, Bratislava, Slovakia. ⁸Australian Centre for Neutron Scattering, Australian Nuclear Science and Technology Organisation, Sydney, NSW, Australia. ⁹Department of Research and Conservation, Allwetterzoo Münster, Münster, Germany. ¹⁰Department of Cell and Developmental Biology, University College London, London, UK.

*Corresponding author. Email: juand.daza@gmail.com

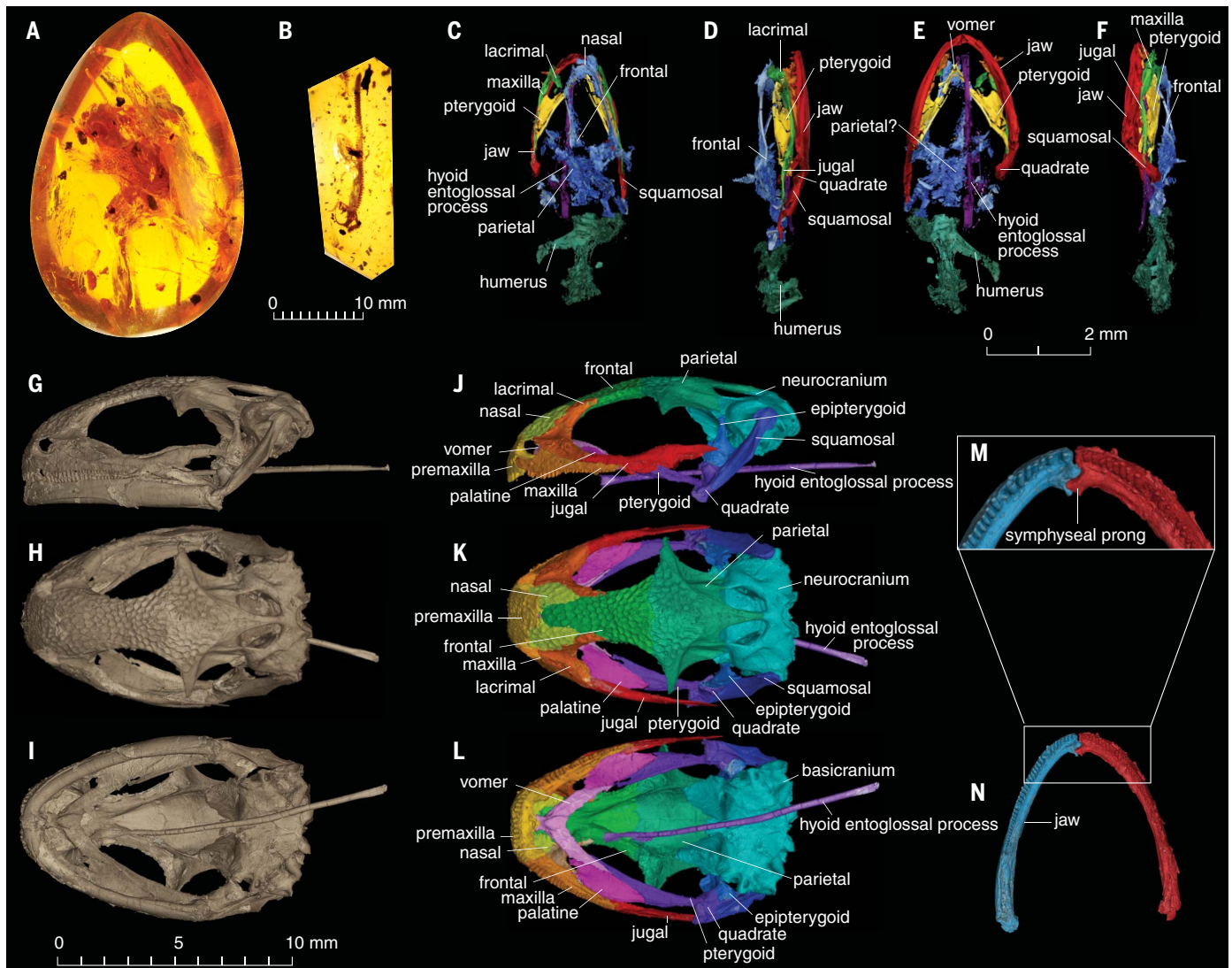


Fig. 1. Holotype and paratype of *Y. perettii*. Holotype GRS-Ref-060829 (A and G to I) and paratype JZC Bu154 (B to F). High-resolution computed tomography (HRCT) of the paratype with segmented bones [(C) to (F)]; the holotype with jaw articulated [(G) to (I)]; the holotype with all bones segmented and jaw removed [(J) to (L)]; and the jaw of the holotype, with close-up of the mandibular symphysis (M and N). Lateral [(D), (F), (G), and (J)], dorsal [(C), (H), (K), (M), and (N)], and ventral [(E), (I), and (L)] views.

most closely resembles the older (120 Ma) Japanese *Shirerpeton* (7), but the phylogenetic analysis placed *Yaksha* as the sister taxon to *Shirerpeton* and either a clade of derived *Albanerpeton* species (Fig. 4A) or *Albanerpeton nexuosum*. Both *Yaksha* and *Shirerpeton* are thus nested within *Albanerpeton* as currently defined. However, we support the view that the single genus *Albanerpeton*, extending from the mid-Cretaceous of North America [~100 Ma (3)] to the Early Pleistocene of Italy [~2 Ma (2)], requires revision and subdivision (7), with the name *Albanerpeton* restricted to the type species, *A. inexpectatum*, and those Cenozoic species that consistently group closely with it (Fig. 4, supplementary text S3.1, and figs. S6 to S11).

As the relationships of albanerpetontids to extant amphibians (Lissamphibia), and the wider relationships of lissamphibians to fossil

amphibians, are currently debated, we updated the coding for albanerpetontids, based on *Yaksha*, into four recent data matrices (20–23). Depending on the matrix used, albanerpetontids were placed (i) as sister to Batrachia (frogs and salamanders) within lepospondyls (20), (ii) as stem-Lissamphibia (22), (iii) nested within Lissamphibia (21), or (iv) nested in a clade comprising derived lepospondyls and lissamphibians (23). Nevertheless, all analyses placed albanerpetontids either as stem- or crown-group Lissamphibia (Fig. 4, C to F). Frustratingly, the new data provided no resolution to the lepospondyl versus temnospondyl origin hypotheses (23, 24) (supplementary text S3.2 and figs. S13 to S16).

Biological implications

The paratype of *Yaksha* (JZC Bu154) was previously identified as a stem-chameleon (14), mainly on the basis of the long entoglossal

hyoid process. In chameleons, this hyoid element is embedded in the base of the tongue, surrounded by collagenous sheaths (25) and a circular accelerator muscle (26). Contraction of the muscle, enhanced by an elastic storage-and-release effect from the wrapping sheaths (25, 27), propels the tongue and entoglossal process out of the mouth, extending the tongue at speeds of up to 500 m/s and to a distance of at least a body length (25). An analogous mechanism exists in some plethodontid salamanders: as the tongue extends, the entire hyoid apparatus is folded into a bundle of cartilaginous rods that is projected out of the mouth again by circular protractor muscles (27, 28). The entoglossal process of *Yaksha* [and other albanerpetontids (1)] is embedded into the remnant of the tongue pad (Fig. 2B), which suggests that albanerpetontids also had a ballistic feeding strategy (fig. S1), plausibly using a

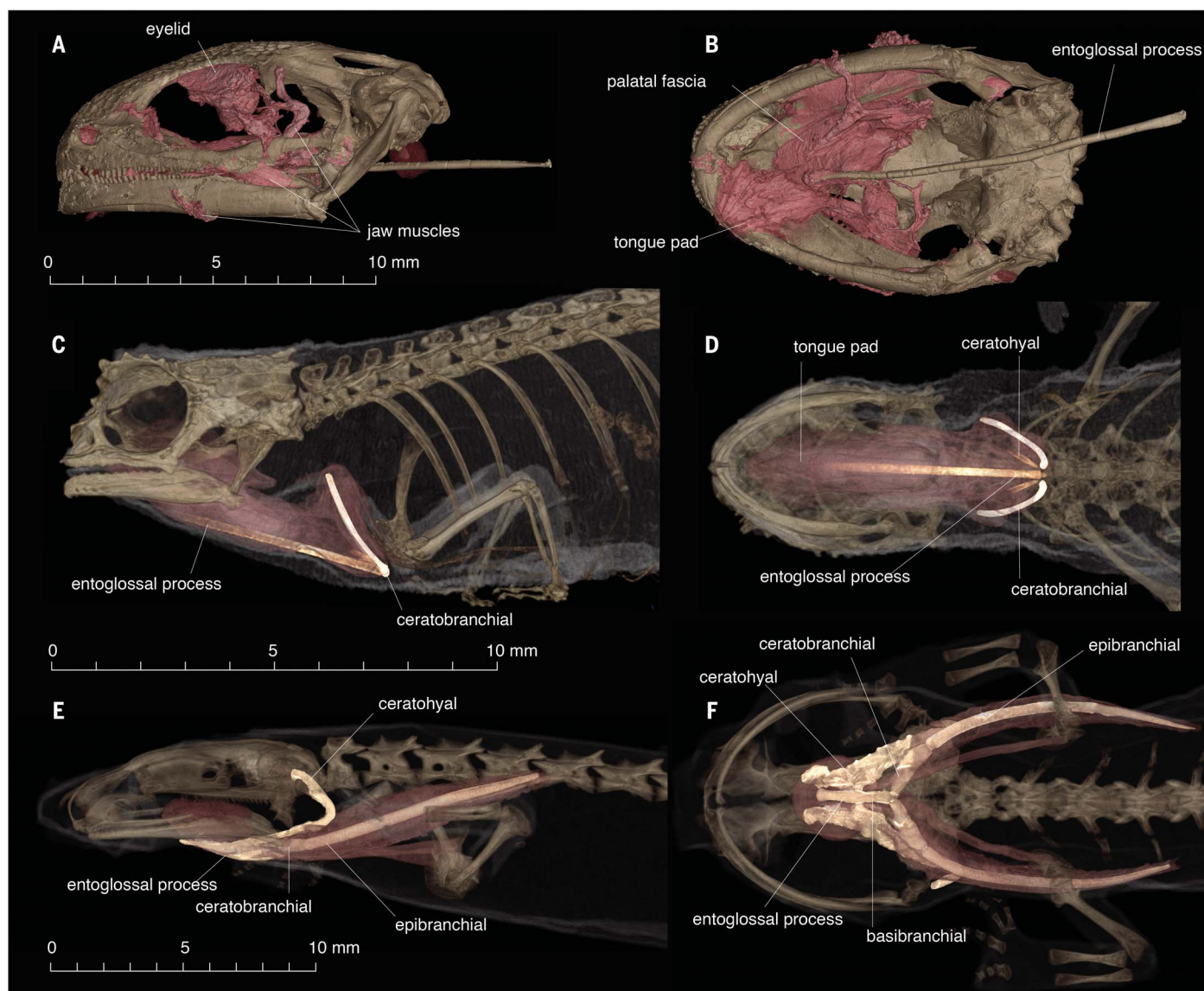


Fig. 2. Comparison of skeletal components in three tetrapods with ballistic tongues. Holotype of *Y. perettii* (GRS-Ref-060829) showing the preserved soft tissue (pink), including the tongue in lateral (A) and ventral (B) views. Diffusible iodine-based contrast-enhanced computer tomographies (DiceCT) of a leaf litter chameleon [*Brookesia* sp. UADBA:herps:15550 (34)] in lateral (C) and ventral (D) views. DiceCT of lungless salamander [*Bolitoglossa porrasorum* UF156522 (35)] in lateral (E) and ventral (F) views.

similar combination of circular muscles and wrapping collagen. This interpretation of *Yaksha* as a chameleon analog may help to explain some of the specialized features of albanerpetontids, including the complex neck and jaw joints, the long curved unguis covered in claw sheaths (fig. S3), and the large forward-looking orbits. We therefore interpret *Yaksha* (and all albanerpetontids) as a sit-and-wait predator, living on or around trees, and using a ballistic tongue to catch small invertebrates. In external appearance, albanerpetontids probably resembled tiny lizards more than salamanders.

There is a corollary to this interpretation of albanerpetontids as ballistic feeders. Most lissamphibians use a combination of lung breathing and cutaneous respiration. Lacking

rib-mediated ventilation, these amphibians inflate their lungs by buccal pumping, using the hyoid apparatus to raise and depress the floor of the mouth (12). Plethodontid salamanders are lungless and rely on cutaneous gas exchange; this freed the hyoid apparatus from its constraints (29), permitting modification for specialized tongue feeding. By analogy, this respiratory mode must also apply to albanerpetontids, given their derived hyoid anatomy. Although the possession of scales might seem contradictory, studies on scaled caecilians have revealed some capacity for cutaneous gas exchange (30). For miniaturized amphibians such as albanerpetontids, it may have been sufficient.

The paratype of *Yaksha* (JZC Bu154) has a skull one-fourth the size of that of the mini-

aturized adult specimen of *Yaksha*. In amphibians, miniaturization is frequently coupled with direct development (31). The diminutive paratype of *Yaksha* may therefore indicate that albanerpetontids lacked a free-living larva.

Yaksha provides considerable new morphological, functional, and phylogenetic data for this enigmatic Jurassic-Pleistocene clade. Albanerpetontids are revealed as specialized, sit-and-wait terrestrial predators, combining lifestyle features of miniaturized chameleons and plethodontid salamanders and extending the history of ballistic tongue feeding by ~100 Ma, given molecular divergence estimates dating the origins of crown chameleons (32) and plethodontids (33) to the Late Cretaceous or Paleocene.

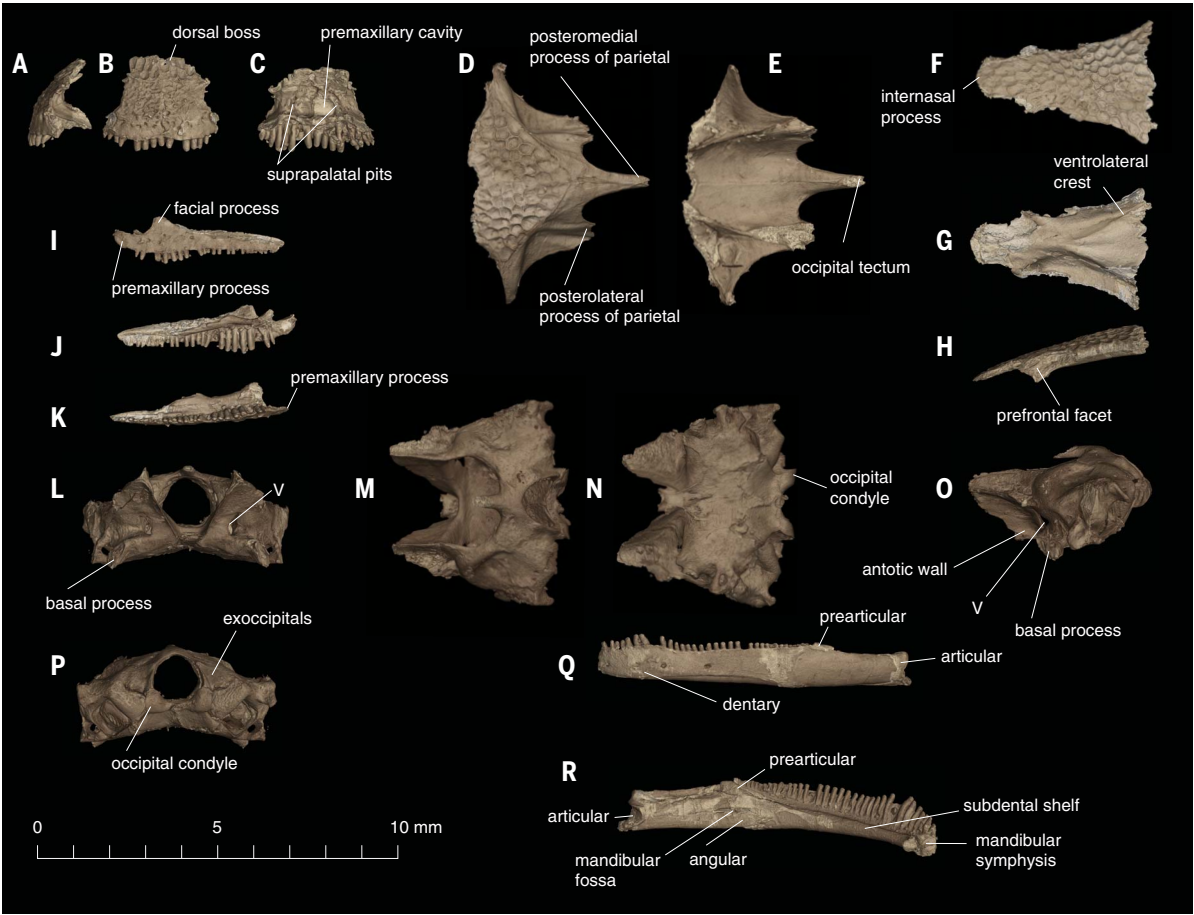
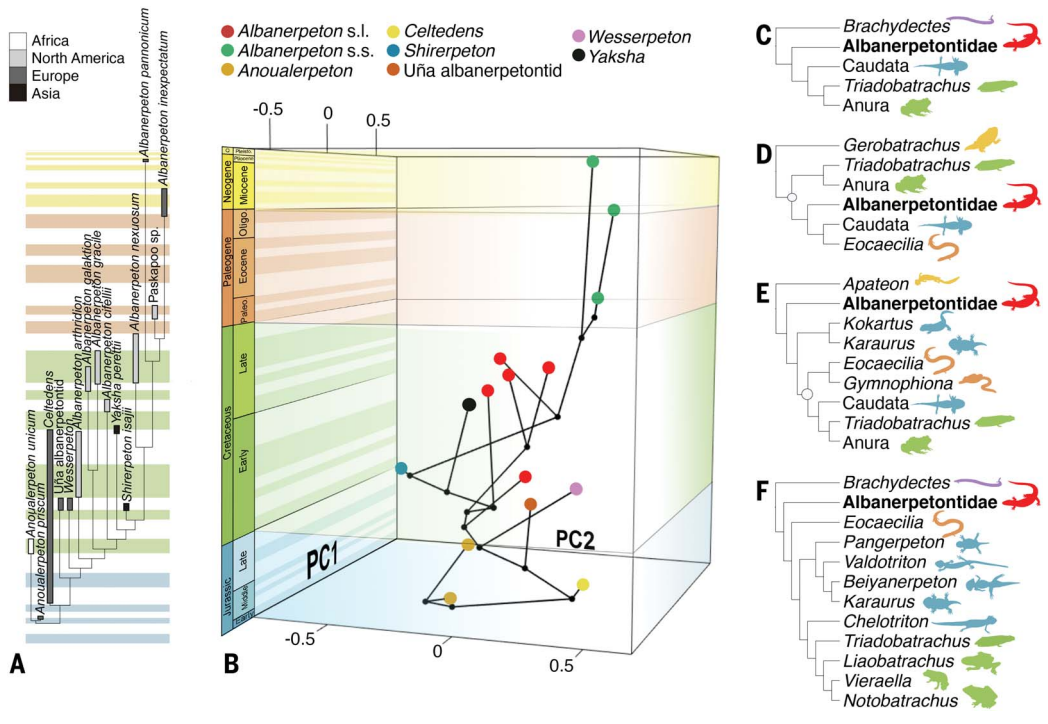


Fig. 3. Isolated elements of the holotype skull of *Y. perettii* (GRS-Ref-060829). Fused (or tightly paired) premaxillae (A to C), parietal (D and E), frontal (F to H), left maxilla (I to K), neurocranium (L to P), and left mandible (Q and R). Lateral [(A), (H), (I), (O), and (Q)], medial [(J) and (R)], anterior [(B) and (L)], posterior [(C) and (P)], dorsal [(D), (F), and (M)], and ventral [(E), (G), (K), and (N)] views.

Fig. 4. Phylogenetic relationships of albanerpetontids calibrated against age and chronophylomorphospace using the dataset for albanerpetontids.

(A) Phylogenetic relationships of albanerpetontids with geographical distribution plotted against time. (B) First and second principal components (PC1 and PC2) explain 35.5 and 23.8% of the total variance, respectively. Note the position of purported members of the genus *Albanerpeton* in separate clades and regions of the morphospace. The tree was dated using the “equal” method, but an alternative tree dated under the “mbi” method is also provided in the supplementary materials. s.l., sensu lato; s.s., sensu stricto. (C to F) Alternative positions recovered for Albanerpetontidae in relation to lissamphibians using different data matrices: (C) (25), (D) (26), (E) (27), and (F) (28).



REFERENCES AND NOTES

1. G. McGowan, *Zool. J. Linn. Soc.* **135**, 1–32 (2002).
2. A. Villa, H. A. Blain, M. Delfino, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **490**, 393–403 (2018).
3. J. D. Gardner, *Palaeontology* **42**, 529–544 (1999).
4. J. D. Gardner, D. G. DeMar Jr., *Palaeobiodivers. Palaeoenvir.* **93**, 459–515 (2013).
5. J. Gardner, M. Böhme, in *Vertebrate Microfossil Assemblages*, J. T. Sankey, S. Baszio, Eds. (Indiana Univ. Press, 2008), pp. 178–218.
6. P. Skutschas, *Acta Palaeontol. Pol.* **52**, 819–821 (2007).
7. R. Matsumoto, S. E. Evans, *PLOS ONE* **13**, e0189767 (2018).
8. J. D. Gardner, S. E. Evans, D. Sigogneau-Russell, *Acta Palaeontol. Pol.* **48**, 301–319 (2003).
9. H. Haddoumi et al., *Gondwana Res.* **29**, 290–319 (2016).
10. G. McGowan, S. Evans, *Nature* **373**, 143–145 (1995).
11. R. Estes, Ed., *Handbuch der Palaoherpologie*, 2. *Gymnophiona, Caudata* (Gustav Fischer Verlag, 1981).
12. W. E. Duellman, L. Trueb, *Biology of Amphibians* (Johns Hopkins Univ. Press, 1994).
13. R. C. Fox, B. G. Naylor, *Can. J. Earth Sci.* **19**, 118–128 (1982).
14. J. D. Daza, E. L. Stanley, P. Wagner, A. M. Bauer, D. A. Grimaldi, *Sci. Adv.* **2**, e1501080 (2016).
15. Peretti Museum Foundation, MicroCT scan of GRS-ref-060829, *Yaksha perettii*, MorphoSource (2020); <https://doi.org/10.17602/M2/M104768>.
16. James Zigras Collection, MicroCT scan of JZC-BU-154, *Yaksha perettii*, MorphoSource (2020); <https://doi.org/10.17602/M2/M126155>.
17. Peretti Museum Foundation, MicroCT scan of GRS-ref-27746, *Yaksha perettii*, MorphoSource (2020); <https://doi.org/10.17602/M2/M104766>.
18. G. H. Shi et al., *Cretac. Res.* **37**, 155–163 (2012).
19. S. C. Sweetman, J. D. Gardner, *Acta Palaeontol. Pol.* **58**, 295–324 (2013).
20. A. Huttenlocker, J. Pardo, B. J. Small, J. Anderson, *J. Vertebr. Paleontol.* **33**, 540–552 (2013).
21. J. D. Pardo, M. Szostakiwskyj, P. E. Ahlberg, J. S. Anderson, *Nature* **546**, 642–645 (2017).
22. J. D. Pardo, B. J. Small, A. K. Huttenlocker, *Proc. Natl. Acad. Sci. U.S.A.* **114**, E5389–E5395 (2017).
23. D. Marjanović, M. Laurin, *PeerJ* **6**, e5565 (2019).
24. J. S. Anderson, R. R. Reisz, D. Scott, N. B. Fröbisch, S. S. Sumida, *Nature* **453**, 515–518 (2008).
25. J. H. de Groot, J. L. van Leeuwen, *Proc. R. Soc. London Ser. B* **271**, 761–770 (2004).
26. J. L. Van Leeuwen, *Philos. Trans. R. Soc. London Ser. B* **352**, 573–589 (1997).
27. A. Sakes et al., *PLOS ONE* **11**, e0158277 (2016).
28. S. M. Deban, D. B. Wake, G. Roth, *Nature* **389**, 27–28 (1997).
29. R. Eric Lombard, D. B. Wake, *J. Morphol.* **153**, 39–79 (1977).
30. A. W. Smits, J. I. Flanagan, *Am. Zool.* **34**, 247–263 (1994).
31. D. L. Levy, R. Heald, *Cold Spring Harbor Perspect. Biol.* **8**, a019166 (2015).
32. F. T. Burbrink et al., *Syst. Biol.* **69**, 502–520 (2020).
33. X. X. Shen et al., *Syst. Biol.* **65**, 66–81 (2016).
34. University of Antananarivo Biological Collections, DICECT scan of UADBA-herps-15550, *Brookesia* sp., MorphoSource (2020); <https://doi.org/10.17602/M2/M163334>.
35. University of Florida, Florida Museum of Natural History, DICECT scan of UF-herp-156522, *Bolitoglossa porrasorum*, MorphoSource (2018); <https://doi.org/10.17602/M2/M43742>.

ACKNOWLEDGMENTS

We thank A. Peretti and the Peretti Museum Foundation for access to their amber collection; GRS staff for organizing the Amber Symposium in Bangkok; J. A. Maisano and M. Colbert (UTCT, University of Texas) for scans; A. Maksimenko for technical assistance at the Australian Synchrotron; D. Grimaldi (AMNH) for paratype photographs and specimen notes; S. Abramowicz (LACM) for reconstruction of *Yaksha*; J. Gardner and an anonymous reviewer for their comments on the manuscript; and A. Raselimanana (U. d'Antananarivo) for access to specimens. **Funding:** National Science Foundation Division of Environmental

Biology (NSF DEB) grant 1657656, Sam Houston State University (J.D.D.); NSF DBI1701714 (E.L.S.); NF170464 (Royal Society, UK), IJC2018-037685-I, the project CGL 2017-82654-P (MICINN, Spain, and FEDER, EU), and the CERCA Programme/Generalitat de Catalunya (A.B.); and grant 1/0209/18 (Ministry of Education of Slovak Republic and Slovak Academy of Sciences) (A.Č.). **Ethics statement:** Specimens were acquired following the ethical guidelines for the use of Burmese amber set forth by the Society for Vertebrate Paleontology (see “Provenance and Ethical Statement” section of the supplementary materials for a detailed description of ethical fossil acquisition and accession). We hope that this study will serve as a model for other researchers working with these types of materials in this region. **Author contributions:** Project was initiated by S.E.E., J.D.D., and E.L.S. E.L.S. and J.J.B. processed HRCT data. J.S.A., J.D.D., and A.B. performed phylogenetic analyses, with coding input from S.E.E. A.B. performed morphospace analyses. The paper was drafted by S.E.E., with contributions from all authors. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are available in the main paper or the supplementary materials. The holotype of *Yaksha perettii* (GRS-Ref-060829) and referred specimen are housed at the Peretti Museum Foundation, Switzerland. The paratype (JZC Bu154) belongs to the James Zigras Collection, housed at the American Museum of Natural History, New York, USA.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6517/687/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S15
References (36–74)

5 March 2020; accepted 8 September 2020
10.1126/science.abb6005

CORONAVIRUS

Epidemiology and transmission dynamics of COVID-19 in two Indian states

Ramanan Laxminarayan^{1,2,3}, Brian Wahl^{3,4}, Shankar Reddy Dudala⁵, K. Gopal⁶, Chandra Mohan B⁷, S. Neelima⁸, K. S. Jawahar Reddy⁹, J. Radhakrishnan¹⁰, Joseph A. Lewnard^{11,12*}

Although most cases of coronavirus disease 2019 (COVID-19) have occurred in low-resource countries, little is known about the epidemiology of the disease in such contexts. Data from the Indian states of Tamil Nadu and Andhra Pradesh provide a detailed view into severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) transmission pathways and mortality in a high-incidence setting. Reported cases and deaths have been concentrated in younger cohorts than would be expected from observations in higher-income countries, even after accounting for demographic differences across settings. Among 575,071 individuals exposed to 84,965 confirmed cases, infection probabilities ranged from 4.7 to 10.7% for low-risk and high-risk contact types, respectively. Same-age contacts were associated with the greatest infection risk. Case fatality ratios spanned 0.05% at ages of 5 to 17 years to 16.6% at ages of 85 years or more. Primary data from low-resource countries are urgently needed to guide control measures.

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the virus that causes coronavirus disease 2019 (COVID-19), has spread rapidly around the world since emerging in Wuhan, China, in late 2019 (1). Our current understanding of COVID-19 comes largely from disease surveillance and epidemiologic studies undertaken during the early phases of the pandemic in China (1–3) and in the high-income countries of Europe (4, 5) and North America (6–8). However, most confirmed cases of COVID-19 have now occurred in low- and middle-income countries (LMICs), where a substantial proportion of individuals may be at increased risk of severe outcomes and face barriers to accessing quality health services (9–11). Although multiple modeling studies have sought to assess how COVID-19 might affect individuals and communities in such settings (12–14), almost no primary studies of the transmission dynamics and clinical outcomes of COVID-19 in LMICs are available to validate these models and inform intervention strategies (15).

More than 1.3 billion people are at risk of SARS-CoV-2 infection in India, where concerns over COVID-19 have prompted large-scale containment strategies at the national, state, and local levels (16). The country's first known COVID-19 case, documented on 30 January 2020, was an Indian national evacuated from China (17). Andhra Pradesh and Tamil Nadu are two states in the south of India whose 127.8 million residents collectively account for about 10% of the country's total population. Although they are not the wealthiest states in India, Andhra Pradesh and Tamil Nadu are among the states with the largest health care workforces and public health expenditures per capita, and are known for their effective primary health care delivery models (18–20). Both states initiated rigorous disease surveillance and contact tracing early in response to the pandemic. Procedures include syndromic surveillance and SARS-CoV-2 testing for all individuals seeking care for severe acute respiratory illness or influenza-like illness at health care facilities; delineation of 5-km “containment zones” surrounding cases for daily house-to-house surveillance to identify individuals with symptoms; and daily follow-up of all contacts of laboratory-confirmed or suspected COVID-19 cases, with the aim of testing these individuals 5 to 14 days after their contact with a primary case, irrespective of symptoms, to identify onward transmission (21, 22). We analyzed comprehensive surveillance and contact-tracing data from these programs in an effort to understand transmission dynamics and clinical outcomes of COVID-19 in South India, and to provide insights into control of SARS-CoV-2 in similar LMIC settings.

Expansion of SARS-CoV-2

In India, surveillance of COVID-19 was initiated with airport screening for severe acute respi-

ratory infection, especially for travelers from China. Tamil Nadu further instituted thermal and clinical screening at land borders with other states on 4 March 2020. Nationwide, testing was initially prioritized for symptomatic individuals with history of travel or contact with a confirmed COVID-19 case within the previous 14 days, and was expanded to include all symptomatic individuals and asymptomatic contacts of confirmed cases in states between 20 and 28 March 2020. We detail the timeline of changes in surveillance practices at federal and state levels in the supplementary materials.

Tamil Nadu and Andhra Pradesh each recorded their first laboratory-confirmed COVID-19 cases on 5 March. Under-ascertainment of cases during March and early April was likely due to limited testing availability and testing algorithms. The proportion of tests yielding positive results peaked at 39.7% in Tamil Nadu and 33.5% in Andhra Pradesh on 30 and 31 March 2020, respectively, when the daily number of tests performed was low in the two states (range, 379 to 469 tests; Fig. 1). Throughout early April, increases in the number of tests performed daily coincided with a reduction in the proportion of tests yielding positive results. Our analyses include data collected through 1 August, at which time Tamil Nadu and Andhra Pradesh had identified 263,330 and 172,209 cases, respectively (table S1). (Because testing and contact tracing constitute routine public health activities, data collection was not governed by an institutional review board.)

The earliest clusters of locally acquired cases emerged in March in Chennai and surrounding coastal districts of eastern Tamil Nadu. Of all districts, Chennai ultimately experienced the highest cumulative incidence of COVID-19, totaling 102,199 cases (204.6 per 10,000 population) by 1 August 2020. An outbreak beginning on 28 April caused 1142 cases by 15 May in the adjoining districts of Ariyalur, Cuddalore, Perambalur, and Villuppuram in Tamil Nadu; thereafter, few cases were identified in these districts until early June (fig. S1). Although limited in March and April, incidence in southern districts of Tamil Nadu surrounding Madurai increased during June and reached rates commensurate with incidence in the northern districts of Chennai, Kancheepuram, and Tiruvallur by 1 August, with one to four new positive detections per 10,000 population daily. Similar increases in incidence occurred throughout all districts of Andhra Pradesh in June, where the numerical and geographic extent of cases remained limited during April and May despite similar levels of testing relative to Tamil Nadu.

Statewide estimates of the time-varying reproduction number R_t , describing the number of secondary infections that each infected individual would be expected to generate (23), declined from a range of 1.7 to 3.0 in Tamil

¹Center for Disease Dynamics, Economics and Policy, New Delhi, India. ²Princeton Environmental Institute, Princeton University, Princeton, NJ, USA. ³Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. ⁴International Vaccine Access Center, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA. ⁵Department of Community Medicine, Government Medical College, Kadapa, Andhra Pradesh, India. ⁶Animal Husbandry, Dairying and Fisheries Department, Government of Tamil Nadu, Chennai, Tamil Nadu, India. ⁷Backward Classes, Most Backward Classes, and Minorities Welfare Department, Government of Tamil Nadu, Chennai, Tamil Nadu, India. ⁸Department of Community Medicine, Guntur Medical College, Guntur, Andhra Pradesh, India. ⁹Department of Health, Family Welfare, and Medical Education, Government of Andhra Pradesh, Amaravati, Andhra Pradesh, India. ¹⁰Health and Family Welfare Department, Government of Tamil Nadu, Chennai, Tamil Nadu, India. ¹¹Division of Epidemiology, School of Public Health, University of California, Berkeley, CA, USA. ¹²Center for Computational Biology, College of Engineering, University of California, Berkeley, CA, USA.

*Corresponding author. Email: jlewnard@berkeley.edu

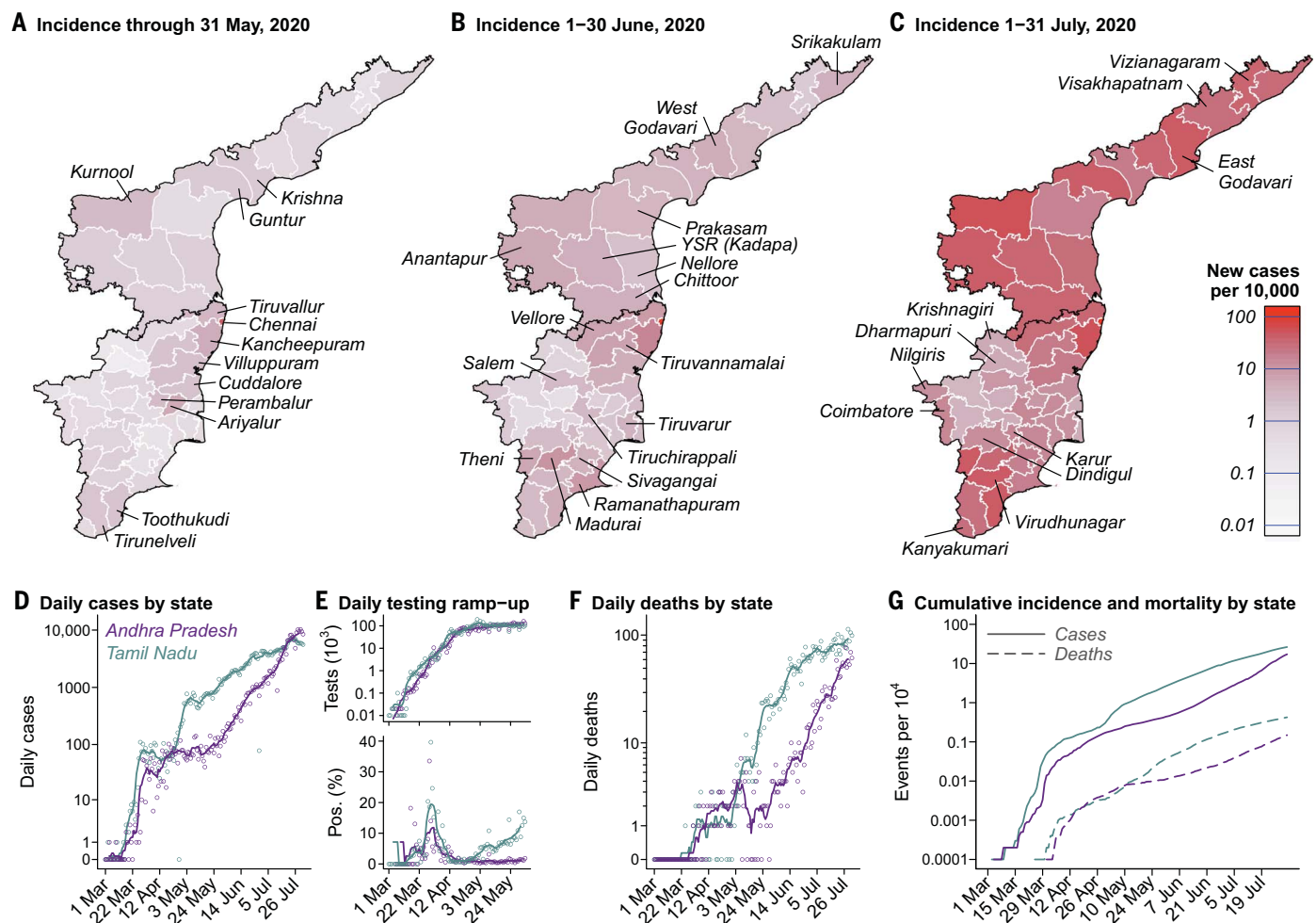


Fig. 1. Incidence over time and across districts in Tamil Nadu and Andhra Pradesh. (A to C) Red shading of regions on the choropleth map indicates higher incidence over each period: 1 March to 31 May 2020 (A), 1 to 30 June 2020 (B), and 1 to 31 July 2020 (C). Districts are plotted according to 2019 administrative boundaries and do not reflect the recent bifurcation of Tirunelveli, Villuppuram, Vellore, and Chengalpattu districts. (D) Cases detected each day in each state (points) and 7-day moving averages (lines). Cases are aggregated by testing date; data are plotted in blue and lavender for Tamil Nadu and Andhra Pradesh, respectively, for all figure panels. (E) Diagnostic tests conducted each day (top)

and the proportion of tests yielding positive results (bottom) for the period March through May 2020, when districts reported comprehensive testing information to the state governments. Points and lines indicate daily counts and 7-day moving averages, respectively. The high proportion of positive tests from late March to mid-April, while case number remained relatively stable, may indicate a period during which cases were undercounted because of limited testing capacity. (F) Daily deaths in the two states. Points and lines indicate daily counts and 7-day moving averages, respectively. (G) Cumulative incidence (solid lines) and mortality (dashed lines) per 10,000 population.

Nadu and 1.4 to 4.3 in Andhra Pradesh over the period 10 to 23 March to a range of 1.0 to 1.3 in both states by the third week of the initial countrywide lockdown (fig. S3). Expansions in testing over this same period, however, are likely to bias analyses of changes in R_t over time (24). Estimates of R_t held in the range 1.1 to 1.4 from 15 May onward within both states, although incidence trajectories differed over time by district (fig. S1), likely reflecting changes in both the uptake and enforcement of social distancing interventions as well as the effectiveness of contact-tracing efforts.

Contact tracing

Contact-tracing efforts in the states reached 3,084,885 known exposed contacts of confirmed

cases by 1 August 2020 (table S2); individual-level epidemiological data on cases and contacts, as well as laboratory test results, were available from 575,071 tested contacts of 84,965 confirmed cases. Traced contacts tended to be younger and were more often female than their linked index cases (table S3). Additionally, test-positive individuals identified through contact tracing were, on average, 1.3 years (bootstrap 95% confidence interval, 1.1 to 1.5 years) younger and 4.5% (3.7 to 5.4%) less likely to be male than the overall population of COVID-19 cases in the two states (table S4). Because studies in other settings have shown the risk of symptomatic disease to be higher among older age groups and among males (25), these findings may indicate the identifi-

cation of less-severe infections through active case finding.

The mean number of contacts tested per index case was 7.3 (interquartile range, 2 to 9), and 0.2% of index cases were linked to >80 tested contacts (range, 1 to 857; Fig. 2A). Numbers of contacts tested varied by district, and the geographic distribution of index cases included in our analyses did not necessarily reflect the geographic distribution of all reported cases (table S5). No positive contacts were identified for 70.7% of index cases for whom reliable contact-tracing data, including test results, were available (Fig. 2A). The distribution of the number of positive contacts linked to each index case was heavily right-skewed, and we estimated a negative binomial dispersion

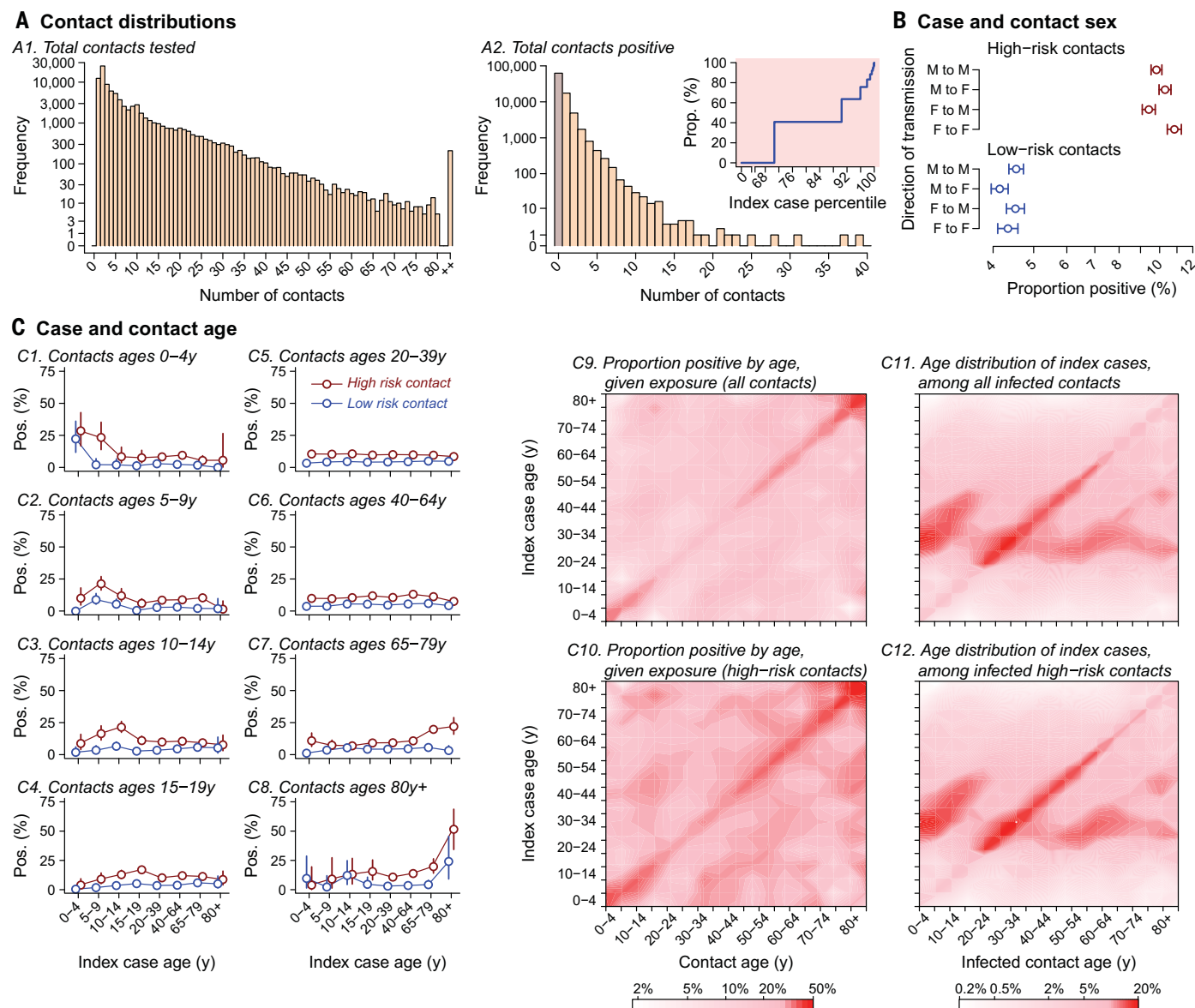


Fig. 2. Analyses of contact-tracing data for 575,071 tested contacts of 84,965 infected individuals from whom test results were available, together with individual-level detailed epidemiological data on exposed contacts and index cases. (A) Left: Distribution of the number of contacts traced for each index case in Tamil Nadu and Andhra Pradesh, binning values ≥ 80 (0.2%). Right: Number of positive contacts traced from each index case. The inset shows the cumulative attributable proportion of secondary infections (y axis) associated with quantiles (x axis) of the distribution of the number of positive contacts traced per index case; percentiles 0 and 100 indicate index cases with the fewest and the most positive contacts identified, respectively. **(B)** Adjusted estimates from Poisson regression models addressing the proportion of female and male contacts with a positive result among those

who were known to be exposed to female and male index cases; models further control for case and contact age groups (interacted) and for state. We stratify for high-risk and low-risk contacts, as defined in table S6. Points and lines indicate mean estimates and 95% confidence intervals. **(C)** Proportion of contacts with a positive test result stratified by case and contact age, for high-risk and low-risk contacts. At right, contour plots indicate the proportion of exposed contacts with a positive test result by case and contact age for all contacts and high-risk contacts on a choropleth scale; see table S8 for raw counts. Positive test results among tested, exposed contacts are interpreted as evidence of probable transmission from the index case. Also plotted are the age distributions of index cases for all infected contacts and for infected high-risk contacts.

parameter for the distribution of the number of infected contacts traced to each index case of 0.51 (95% confidence interval, 0.49 to 0.52). On average, 9.2 contacts were tested for each index case with ≥ 1 contact identified, as compared to 5.7 contacts tested for each index case without positive contacts identified

(two-sided bootstrap $P < 0.001$; fig. S4). Although our analysis is limited in that it does not necessarily capture all secondary infections (e.g., among contacts who were not reported), these observations are consistent with the presence of superspreading related to differences in individual contact patterns (26).

Assuming that test-positive contacts were infected by the index case to whom they were traced, we estimated that the overall secondary attack rate (or risk of transmission from an index case to an exposed contact) was 10.7% (10.5 to 10.9%) for high-risk contacts, who had close social contact or direct physical

contact with index cases without protective measures, and 4.7% (4.6 to 4.8%) for low-risk contacts, who were in the proximity of index cases but did not meet these criteria for high-risk exposure (tables S6 and S7). Data on exposure settings, available for 18,485 contacts of 1343 index cases, revealed considerable differences in transmission risk associated with differing types of interaction. Secondary attack rate estimates ranged from 1.2% (0.0 to 5.1%) in health care settings to 2.6% (1.6 to 3.9%) in the community and 9.0% (7.5 to 10.5%) in the household. Among 78 individuals with high-risk travel exposures—defined as close proximity to an infected individual in a shared conveyance for ≥ 6 hours—we estimated a secondary attack rate of 79.3% (52.9 to 97.0%).

Whereas secondary attack rate estimates did not differ considerably with respect to the sex of cases and their contacts (Fig. 2B), analyses stratified by case and contact age identified the highest probability of transmission, given exposure, within case-contact pairs of similar age (Fig. 2C and table S8). These patterns of enhanced transmission risk in similar-age pairs were strongest among children aged 0 to 14 years and among adults aged ≥ 65 years, and may reflect differences in the nature of intragenerational and intergenerational social and physical interactions in India (27). Nonetheless, the greatest proportion of test-positive contacts within most age groups were exposed to index cases aged 20 to 44 years (Fig. 2C, fig. S5, and table S8). Serological surveys in other settings have demonstrated that case-based surveillance may lead to underestimation of SARS-CoV-2 infection prevalence among children (28, 29); therefore, it remains crucial to establish whether the role of children in transmission is underestimated in studies such as ours using case-based surveillance to identify index infections.

Mortality among COVID-19 cases

In a subcohort of 102,569 cases in Tamil Nadu and 22,315 cases in Andhra Pradesh who tested positive at least 30 days before the end of the study follow-up period, the overall case fatality ratio was 2.06% (1.98 to 2.14%; Fig. 3). Age-specific estimates ranged from 0.05% (0.012 to 0.11%) at ages 5 to 17 years to 16.6% (13.4 to 19.9%) at ages ≥ 85 years. Risk of death was higher among male cases than among female cases overall, and the magnitude of this difference widened in the oldest age groups. Higher mortality in older age groups and among males has similarly been observed in high-income settings (1–7, 30–32).

Half of the cases ascertained before death in Tamil Nadu and Andhra Pradesh succumbed within ≤ 6 days of testing (interquartile range, 3 to 12 days), and 1042 fatal cases (18.2% of 5733 observed) were identified either ≤ 24 hours before death or posthumously. Our estimates

of time to death in Tamil Nadu and Andhra Pradesh are below what has been observed internationally: In the United States, median time to death from the date of hospital admission was 13 days (8), and the World Health Organization estimated that time to death after onset of symptoms could range from 2 to 8 weeks on the basis of data from China (33). Our observations likely indicate that a substantial proportion of patients in Tamil Nadu and Andhra Pradesh are diagnosed late in their disease course, although differences in patients' health status, health care systems capacity, and approaches to end-of-life care may also contribute to variation in time to death.

In a survival analysis of the full cohort, mortality by 1 August 2020 was independently associated with older age, with stepwise increases in the adjusted hazard ratio of time to death for each successive age group besides children aged 0 to 4 years, consistent with our estimates of the case fatality ratio (Fig. 3). Additional predictors of mortality included being male [adjusted hazard ratio, 1.62 (1.52 to 1.73) compared with being female], receipt of a test early in the epidemic [0.87 (0.72 to 1.07) for being tested between 1 May and 30 June, 0.74 (0.61 to 0.91) for being tested between 1 July and 1 August, both relative to testing between 1 March and 30 April], and state of residence [1.08 (1.01 to 1.16) for residents of Tamil Nadu compared with those in Andhra Pradesh].

Among decedents in the two Indian states, the most prevalent comorbid conditions were diabetes (45.0%), sustained hypertension (36.2%), coronary artery disease (12.3%), and renal disease (8.2%; table S9). Although prevalence of any comorbidity was highest among decedents at older ages, this pattern differed across conditions; diabetes was most prevalent among decedents aged 50 to 64 years, and liver disease and renal disease were most prevalent in fatal cases at ages 0 to 17 years and 18 to 29 years, respectively. At least one comorbid condition was noted among 62.5% of fatalities, in comparison to 22% of fatalities in the United States as of 30 May 2020 (34).

Epidemiological comparison to high-income settings

Cases in Tamil Nadu and Andhra Pradesh showed a younger age distribution than cases reported in the United States as of 21 August 2020 (Fig. 4) (35). Comparison of cumulative COVID-19 incidence across ages showed that the observed differences surpassed expectations based on population age distributions alone, as signaled by the absence of parallel trends in age-specific incidence (table S10). Although lower across all age groups in Tamil Nadu and Andhra Pradesh in comparison to the United States, age-specific COVID-19 incidence increased sharply in both settings between the 5- to 17-year and the 18- to 29-year age groups.

Whereas incidence declined steadily at ages older than 30 to 39 years in the two Indian states, incidence increased at ages of ≥ 65 years in the United States.

In the two Indian states, only 17.9% of COVID-19 deaths occurring on or before 1 August 2020 were among individuals aged ≥ 75 years, compared with 58.1% of COVID-19 deaths in the United States (Fig. 4 and table S10). Age-specific COVID-19 mortality was lower in Tamil Nadu and Andhra Pradesh than in the United States, consistent with the lower reported incidence of disease. Although COVID-19 mortality trended upward across ages in the two Indian states, mortality plateaued at ages of ≥ 65 years, in contrast to observations in the United States where COVID-19 mortality reached 69.6 deaths per 10,000 individuals aged ≥ 85 years; this observation was consistent with the relatively lower incidence of disease at the oldest ages within the two Indian states.

Discussion

Our findings, based on comprehensive surveillance and contact-tracing data from the Indian states of Tamil Nadu and Andhra Pradesh, provide insight into the epidemiology of COVID-19 in resource-limited populations. Our analysis suggests substantial variation in individuals' likelihood of transmitting: No secondary infections were linked to 71% of cases whose contacts were traced and tested. Although the role of children in transmission has been debated (36, 37), we identify high prevalence of infection among children who were contacts of cases around their own age; this finding of enhanced infection risk among individuals exposed to similar-age cases was also apparent among adults. School closures and other nonpharmaceutical interventions during the study period may have contributed to reductions in contact among children. Nonetheless, our analyses suggest that social interactions among children may be conducive to transmission in this setting. Last, our analyses of fatal outcomes reveal an overall case fatality ratio of 2.1%. Even though our estimates of age-specific case fatality ratios are similar to those in other settings, such comparisons are limited by uncertainty in the proportion of infections ascertained as cases (30, 38). Lower relative incidence of COVID-19 among older adults in Tamil Nadu and Andhra Pradesh has contributed to stark differences in the overall case fatality ratio and age distribution of decedents relative to observations in the United States and other high-income countries (32).

Several factors may contribute to our observation of limited COVID-19 incidence and mortality among older adults in Tamil Nadu and Andhra Pradesh. Imperfect surveillance systems may have contributed to under-ascertainment of cases among older adults, although this circumstance is unexpected given strong public

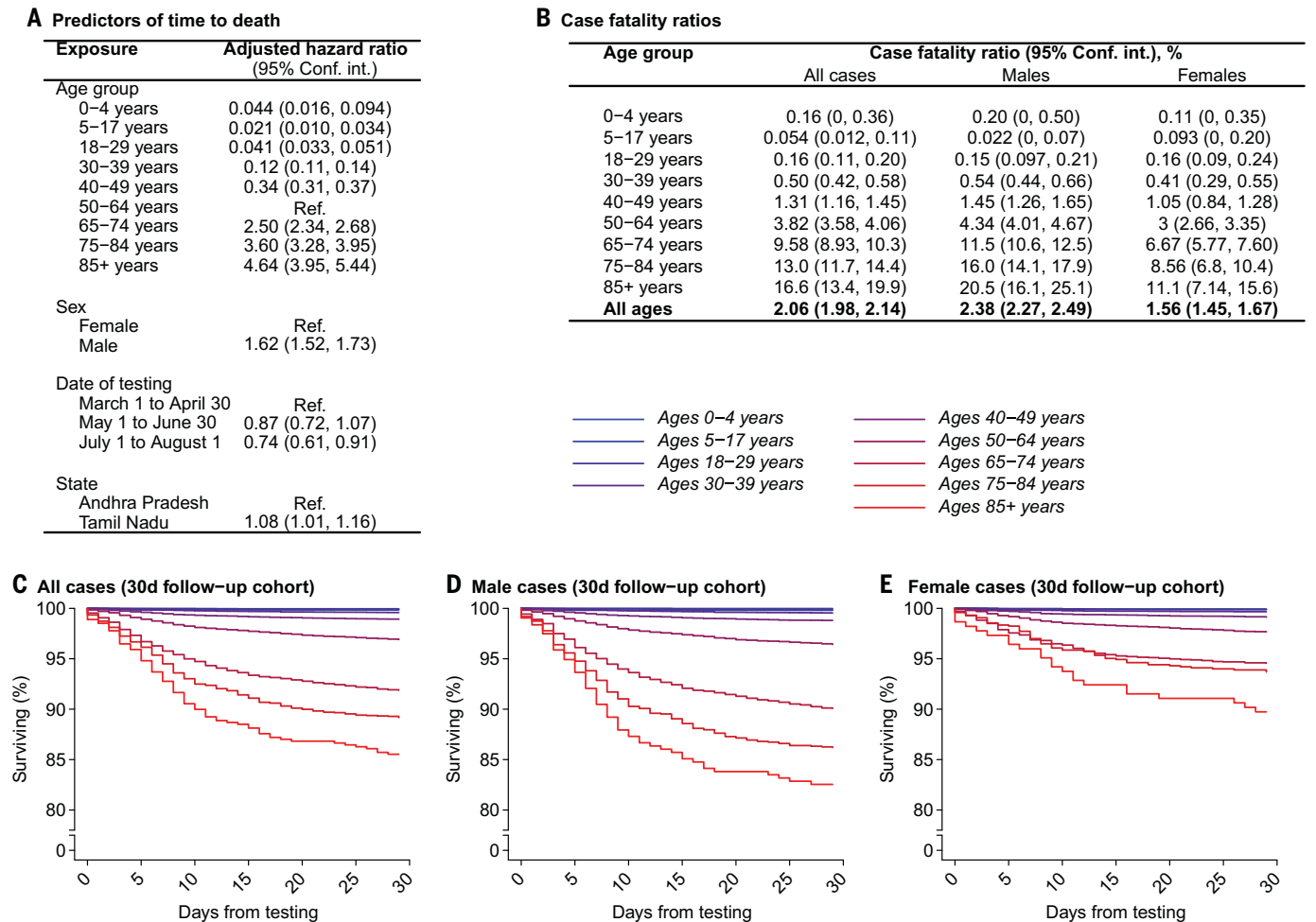


Fig. 3. Mortality among confirmed COVID-19 cases. (A) Adjusted hazard ratios for mortality by 1 August 2020 estimated via Cox proportional hazards models including all confirmed cases. Exposures designated “Ref.” indicate the referent group for hazard ratio calculation. (B) Absolute case fatality risk estimates obtained via bootstrap resampling of individuals with confirmed infection by 1 July 2020. (C to E) Survival probabilities by age within this cohort over the 30-day period after testing, plotted for all cases (C), male cases (D), and female cases (E). Blue-to-red coloration aligns with younger-to-older age group, for strata as defined in the above tables. Age bins were selected on the basis of reporting of U.S. COVID-19 surveillance data (Fig. 4).

and clinical awareness of COVID-19 and the predisposition of older adults to severe disease. Case-based surveillance may likewise underestimate attack rates among younger adult age groups in high-income settings (28, 29). It is plausible that stringent stay-at-home orders for older Indian adults, coupled with delivery of essentials through social welfare programs and regular community health worker interactions, contributed to lower exposure to infection within this age group in Tamil Nadu and Andhra Pradesh. Our finding may also reflect survivorship bias if older adults in India are at disproportionately low risk for SARS-CoV-2 infection relative to the general population—for instance, as a result of higher socioeconomic status (39). Life expectancy at birth is 69 years in India, versus 77 years in China, 79 years in the United States, and 83 years in Italy and South Korea (40); as such, socioeconomic factors distinguishing individuals who survive to

old age from the general population are likely more pronounced in India than in higher-income settings with longer average life expectancies (41, 42). Prospective testing of a large sample of exposed individuals through integrated active surveillance and public health interventions in Tamil Nadu and Andhra Pradesh provided an opportunity to characterize secondary attack rates as a function of both case and contact age, identify risk factors for transmission, and account for deaths outside of health care settings—a limitation of mortality surveillance in other settings (30, 43, 44). However, several limitations should be considered. The contact-tracing data analyzed included only 20% of all reported cases as index cases and represented only 19% of all contacts traced; case-finding efforts further varied by district and over time within Tamil Nadu and Andhra Pradesh. Contacts who complete testing and supply per-

sonal information to tracing teams may not have been representative of the full population. Another limitation was the lack of data on timing of exposure and onset of symptoms in relation to testing dates; this necessitated assumptions about the identification of true index cases. More robust temporal data would reduce the dependence on such assumptions, provide greater insight into the directionality of transmission, and reduce risk for misclassification of infection status among contacts with positive or negative results at the time of testing (45, 46). The lack of temporal data also prevented us from estimating several epidemiologic parameters of interest. Current estimates of both the incubation period (~4 to 6 days) and the serial interval (~3 to 5 days) come from China (1, 47–51). Several factors can modify the incubation period of respiratory viral infections, including the route of acquisition, the infectious dose, and the period

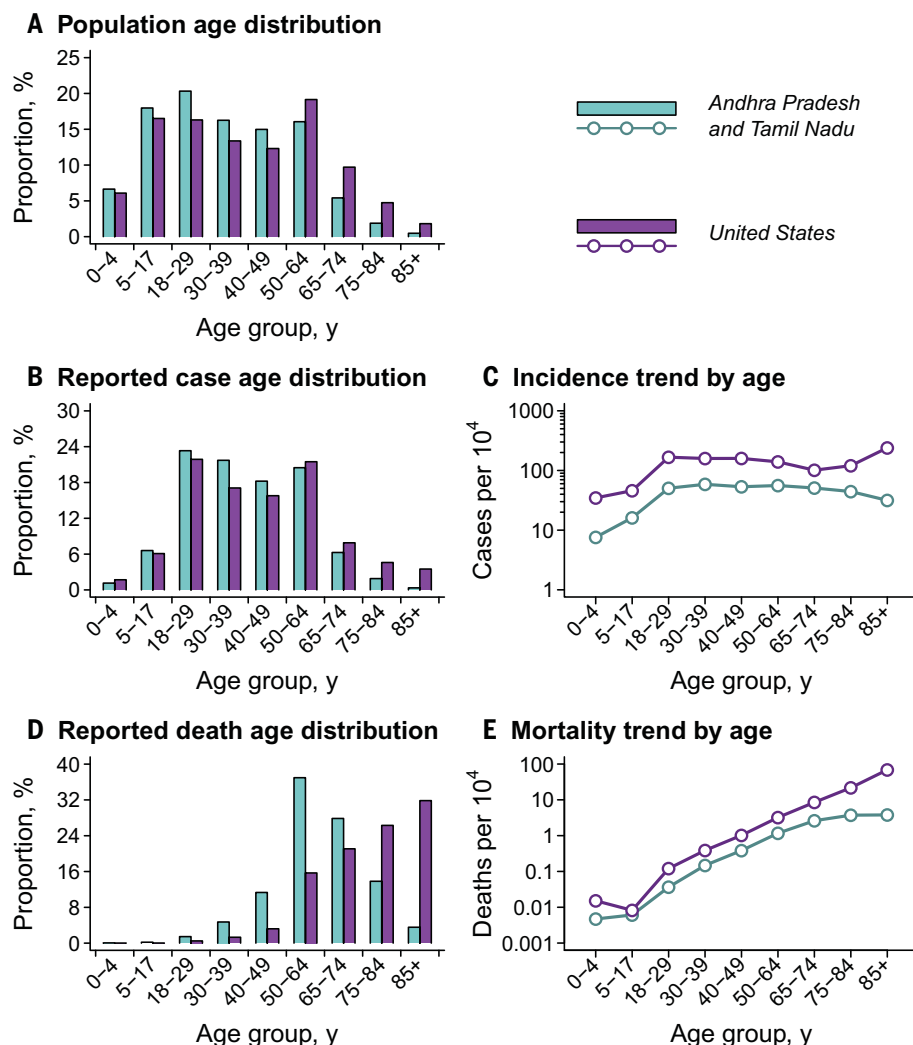


Fig. 4. Demographic comparison of populations, cases, and deaths for Tamil Nadu and Andhra Pradesh versus the United States. (A) Age distribution of the population of Tamil Nadu and Andhra Pradesh (blue) against that of the U.S. population (purple) for comparison; underlying data are shown in table S10. Estimates are census extrapolations for the year 2020 in both settings. (B) Age distribution of cases. (C) Cumulative incidence of COVID-19 by age. (D) Age distribution of deaths. (E) Cumulative COVID-19 mortality by age. Data for the United States include all cases and deaths reported by 21 August 2020 (35).

of exposure to infected cases (52). The serial interval between successive infections is expected to be lower in high-transmission settings. Data allowing estimation of these parameters for SARS-CoV-2 in LMICs are needed to inform quarantine policies and other epidemic response efforts. Some true positives might have been misclassified as a result of imperfect test sensitivity, particularly among contacts tested as few as 5 days after exposure to a confirmed case. Imperfect test sensitivity has been attributed to inadequate sample collection procedures and low viral load in the upper respiratory tract, particularly for pre-symptomatic or asymptomatic cases (53). This limitation could lead to an overall underestimate of transmission risk within case-contact pairs. Finally, although comorbidity data col-

lected as part of COVID-19 mortality surveillance revealed clinical and epidemiological attributes of fatal cases, the fact that such data were not collected for all diagnosed cases prevented inference of the contribution of comorbidities to fatal outcomes.

Surveillance and contact tracing are critical components of an effective public health response to COVID-19 (54, 55). In our study, data generated by these activities within two states of South India provided key insights into the local epidemiology and transmission dynamics of SARS-CoV-2 without competing with emergency response activities for limited resources—a high priority in many LMICs where health workers and diagnostic equipment are already in short supply (15). Similar studies are necessary to inform the

successful adoption of epidemic control measures in low-resource settings globally.

REFERENCES AND NOTES

- Q. Li et al., *N. Engl. J. Med.* **382**, 1199–1207 (2020).
- F. Zhou et al., *Lancet* **395**, 1054–1062 (2020).
- W. J. Guan et al., *N. Engl. J. Med.* **382**, 1708–1720 (2020).
- G. Grasselli et al., *JAMA* **323**, 1574–1581 (2020).
- A. B. Docherty et al., *BMJ* **369**, m1985 (2020).
- S. Richardson et al., *JAMA* **323**, 2052–2059 (2020).
- C. M. Petrilli et al., *BMJ* **369**, m1966 (2020).
- J. A. Lewnard et al., *BMJ* **369**, m1923 (2020).
- E. Dong, H. Du, L. Gardner, *Lancet Infect. Dis.* **20**, 533–534 (2020).
- A. Clark et al., *Lancet Glob. Health* **8**, e1003–e1017 (2020).
- “Coronavirus disease (COVID-19) situation report 209” (World Health Organization, 16 August 2020); www.who.int/docs/default-source/coronaviruse/situation-reports/20200816-covid-19-sitrep-209.pdf?sfvrsn=5dde1ca2_2.
- P. G. T. Walker et al., *Science* **369**, 413–422 (2020).
- M. Gilbert et al., *Lancet* **395**, 871–877 (2020).
- N. G. Davies et al., *Nat. Med.* **26**, 1205–1211 (2020).
- M. Gupta et al., *Glob. Health Res. Policy* **5**, 33 (2020).
- V. Chandrasekhar, “1.3 billion people. A 21-day lockdown. Can India curb the coronavirus?” *Science* 10.1126/science.abc0030 (31 March 2020).
- S. S. Gunthe, S. S. Patra, *Global. Health* **16**, 45 (2020).
- R. Parthasarathi, S. P. Sinha, *Indian J. Community Med.* **41**, 302–304 (2016).
- Planning Commission, “Report of Working Group on National Rural Livelihoods Mission” (Government of India, 2011); <https://ajajeevika.gov.in/en/content/report-working-group-nrlm---2011>.
- A. Pandey, A. P. Galvani, *Lancet Child Adolesc. Health* **4**, 643–645 (2020).
- “Outbreak of coronavirus (COVID-19) disease in China and other countries: Containment plan” (Government of Tamil Nadu, 2020); https://stopcorona.tn.gov.in/wp-content/uploads/2020/03/Corona_Containment_Plan-1.pdf.
- “Prevention and Management of COVID-19: Certain instructions for sample collection, transportation, and testing” (Government of Andhra Pradesh, 2020); [http://hmfw.ap.gov.in/COVID-19/IC/4.GOI Guidelines and Advisories/InstantOrders/COVID Instant Order-11.pdf](http://hmfw.ap.gov.in/COVID-19/IC/4.GOI%20Guidelines%20and%20Advisories/InstantOrders/COVID%20Instant%20Order-11.pdf).
- A. Cori, N. M. Ferguson, C. Fraser, S. Cauchemez, *Am. J. Epidemiol.* **178**, 1505–1512 (2013).
- V. Pitzer et al., The impact of changes in diagnostic testing practices on estimates of COVID-19 transmission in the United States. *medRxiv* 202007338 [preprint]. 24 April 2020.
- R. E. Jordan, P. Adab, K. K. Cheng, *BMJ* **368**, m1198 (2020).
- J. O. Lloyd-Smith, S. J. Schreiber, P. E. Kopp, W. M. Getz, *Nature* **438**, 355–359 (2005).
- S. Kumar et al., *PLOS ONE* **13**, e0209039 (2018).
- M. Pollán et al., *Lancet* **396**, 535–544 (2020).
- F. P. Havers et al., *JAMA Intern. Med.* 10.1001/jamainternmed.2020.4130 (2020).
- G. Onder, G. Rezza, S. Brusaferro, *JAMA* **323**, 1775–1776 (2020).
- H. Salje et al., *Science* **369**, 208–211 (2020).
- N. Sudharsanan, O. Didzun, T. Bärnighausen, P. Geldsetzer, *Ann. Intern. Med.* **M20-2973** (2020).
- “Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19)” (World Health Organization, 2020); www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf.
- E. K. Stokes et al., *MMWR* **69**, 759–765 (2020).
- U.S. Centers for Disease Control and Prevention, “Provisional Death Counts for Coronavirus Disease 2019 (COVID-19);” www.cdc.gov/nchs/nvss/vsrr/covid_weekly/index.htm.
- A. A. Kelvin, S. Halperin, *Lancet Infect. Dis.* **20**, 633–634 (2020).
- B. Lee, W. V. Raszka Jr., *Pediatrics* **146**, e2020004879 (2020).
- R. Verity et al., *Lancet Infect. Dis.* **20**, 669–677 (2020).
- R. Y. N. Chung, D. Dong, M. M. Li, *BMJ* **369**, m1329 (2020).
- “Life expectancy at birth” (World Bank, 2020); doi: 10.1787/how_life-table168n.
- A. S. Deaton, C. H. Paxson, *Am. Econ. Rev.* **88**, 248–253 (1998).
- M. Asaria et al., *BMJ Glob. Health* **4**, e001445 (2019).
- N. P. Jewell, J. A. Lewnard, B. L. Jewell, *JAMA* **323**, 1893–1894 (2020).
- N. P. Jewell, J. A. Lewnard, B. L. Jewell, *Ann. Intern. Med.* **173**, 226–227 (2020).
- Q.-L. Jing et al., *Lancet Infect. Dis.* **20**, 1141–1150 (2020).

46. Y. Yang, I. M. Longini Jr., M. E. Halloran, V. Obenchain, *Biometrics* **68**, 1238–1249 (2012).
47. S. A. Lauer et al., *Ann. Intern. Med.* **172**, 577–582 (2020).
48. J. Zhang et al., *Lancet Infect. Dis.* **20**, 793–802 (2020).
49. Q. Bi et al., *Lancet Infect. Dis.* **20**, 911–919 (2020).
50. Z. Du et al., *Emerg. Infect. Dis.* **26**, 1341–1343 (2020).
51. L. C. Tindale et al., *eLife* **9**, e57149 (2020).
52. J. Lessler et al., *Lancet Infect. Dis.* **9**, 291–300 (2009).
53. S. Woloshin, N. Patel, A. S. Kesselheim, *N. Engl. J. Med.* **383**, e38 (2020).
54. A. S. Fauci, H. C. Lane, R. R. Redfield, *N. Engl. J. Med.* **382**, 1268–1269 (2020).
55. A. J. Kucharski et al., *Lancet Infect. Dis.* **20**, 1151–1160 (2020).
56. De-identified data and code for replication of the analyses: <https://doi.org/10.5281/zenodo.4003365>.

ACKNOWLEDGMENTS

We are indebted to the work of the Governments of Tamil Nadu and Andhra Pradesh as well as health care workers and field workers engaged in outbreak response in these settings. Permission for analysis and publication of the data included in this report was granted by the Governments of Tamil Nadu and Andhra Pradesh.

Funding: J.A.L. received support from the Berkeley Population Center (National Institute of Child Health and Human Development grant P2CHD073964). R.L. received support from NSF grant CCF-1918628 to the Center for Disease Dynamics, Economics & Policy, and from U.S. Centers for Disease Control and Prevention grant 16IPA16092427 to Princeton University. **Author contributions:** Conceptualization, R.L., B.W., J.A.L.; methodology, B.W., J.A.L.; software, J.A.L.; formal analysis, J.A.L.; investigation, S.R.D., K.G., C.M.B., S.N., K.S.J.R., J.R.; resources, S.R.D., K.G., C.M.B., S.N., K.S.J.R., J.R.; data curation, R.L., J.A.L.; writing—original draft, B.W., J.A.L.; writing—review and editing, R.L., B.W., S.R.D., K.G., C.M.B., S.N., K.S.J.R., J.R., J.A.L.; visualization, J.A.L. **Competing interests:** K.G. is the principal secretary to the Government of Tamil Nadu for the Animal Husbandry, Dairying and Fisheries Department. C.M.B. is the principal secretary to the Government of Tamil Nadu for the Department of Backward Classes, Most Backward Classes, and Minorities Welfare. K.S.J.R. is the special chief secretary to the Government of Andhra Pradesh for the Department of Health, Family Welfare, and Medical Education. J.R. is the principal secretary to the Government of Tamil Nadu for the Department of Health and Family. K.G., C.M.B., and J.R. are members of the Team for Epidemic Monitoring, Interventions and Standardizing Health Care Protocols, Government of Tamil Nadu.

All other authors declare no competing interests. **Data and materials availability:** De-identified data and code for replication of the analyses are available at the corresponding author's GitHub page, <https://github.com/joeleward/covid-india> (56). 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/370/6517/691/suppl/DC1
Materials and Methods
Figs. S1 to S5
Tables S1 to S10
MDAR Reproducibility Checklist

11 July 2020; accepted 23 September 2020
Published online 30 September 2020
10.1126/science.abd7672

REPORT

ORGANIC DEVICES

Large-area low-noise flexible organic photodiodes for detecting faint visible light

Canek Fuentes-Hernandez*, Wen-Fang Chou, Talha M. Khan, Larissa Diniz, Julia Lukens, Felipe A. Larrain, Victor A. Rodriguez-Toro, Bernard Kippelen*

Silicon photodiodes are the foundation of light-detection technology; yet their rigid structure and limited area scaling at low cost hamper their use in several emerging applications. A detailed methodology for the characterization of organic photodiodes based on polymeric bulk heterojunctions reveals the influence that charge-collecting electrodes have on the electronic noise at low frequency. The performance of optimized organic photodiodes is found to rival that of low-noise silicon photodiodes in all metrics within the visible spectral range, except response time, which is still video-rate compatible. Solution-processed organic photodiodes offer several design opportunities exemplified in a biometric monitoring application that uses ring-shaped, large-area, flexible, organic photodiodes with silicon-level performance.

Emerging applications such as context-aware optical sensing, touchless interactive computing, wearable biometric monitoring, biomedical imaging, distributed ionizing-radiation detection, and others (1, 2) would benefit from photodetectors with mechanical properties and form factors (e.g., shapes and sizes) that are not physically possible or economically viable with silicon photodiodes (SiPDs). SiPDs remain the photodetector technology of choice for detecting ultraviolet, visible, or near-infrared light because of their high performance and relatively low cost. Noise equivalent power (NEP), defined as the optical power (ϕ_{opt}) producing a signal-to-noise ratio (SNR) of one, is the key measurable metric that quantifies

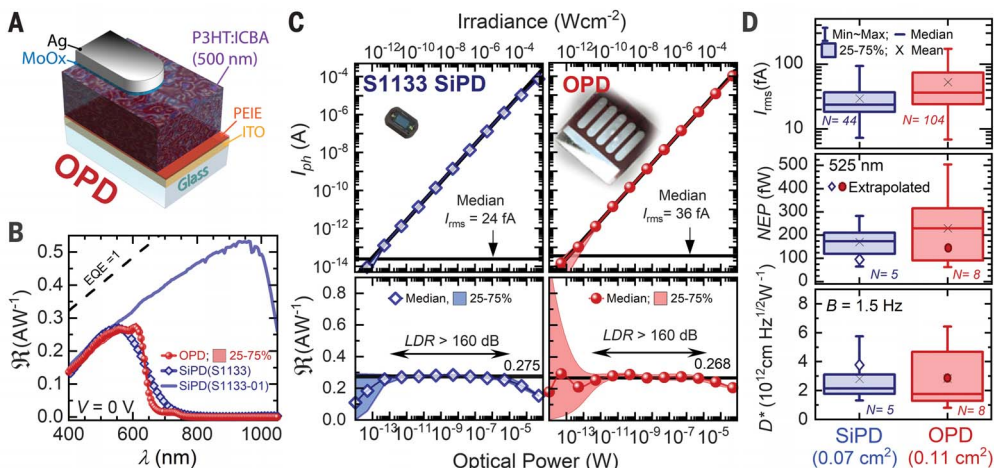
the performance of a photodiode (PD). Because a PD's performance varies with its area (A_{PD}) and measurement bandwidth (B), the specific detectivity, $D^* \equiv \sqrt{A_{\text{PD}}B}/\text{NEP}$, is used to compare performance. Higher D^* values indicate an increased ability to detect fainter light.

Although small-area low-noise SiPDs (e.g., Hamamatsu S1133-01) exhibit white noise-limited specific detectivity (D^*_{wn}) values of $\sim 10^{13} \text{ cm} \cdot \text{Hz}^{1/2} \text{ W}^{-1}$ (NEP_{wn} of $\sim 10 \text{ fW}$) in the visible and near-infrared spectral ranges, area scaling (e.g., $>0.1 \text{ cm}^2$) requires stringent manufacturing control of crystal defects to maintain high D^* values over increasingly large areas (3). To overcome existing trade-offs between performance, area, shape, circuit complexity, power consumption, and cost of photodetectors, PDs using organic molecules (2, 4), metal-halide perovskites (4), and inorganic nanocrystals and quantum dots (4) are being investigated.

Emerging PDs with extrapolated D^* values similar to those of SiPDs have been reported (4, 5), but their accuracy is in question because the NEP is seldom measured and the validity of approximations made to extrapolate D^* values is often overlooked (6). This has hindered progress, given the lack of clarity about the physical principles guiding PD optimization, and makes a comparison of their performance challenging. For instance, because the root mean square electronic noise current (I_{rms}) is seldom measured, injection-blocking layers (2, 4, 7, 8), geometry (i.e., increase photoactive layer thickness) (9–11), and morphological optimization (12) have been studied in the past to reduce the dark current density (J_{dark}) of organic PDs (OPDs) because it was assumed to be proportional to I_{rms} . Despite reports of D^* values of $\sim 10^{12} \text{ cm} \cdot \text{Hz}^{1/2} \text{ W}^{-1}$ or larger, J_{dark} values in OPDs remain in the range of hundreds of picoamperes per square centimeter, and their shunt resistance (R_{p}) values in the range of a few gigohms (8, 10, 11, 13), still orders of magnitude away from values found in low-noise SiPDs (e.g., J_{dark} is a few picoamperes per square centimeter, and R_{p} is hundreds of gigohms; Hamamatsu S1133-series). Despite considerable advances in material development, OPDs that use an injection-blocking layer and a poly(3-hexylthiophene) (P3HT):indene- C_{60} bis-adduct (ICBA) photoactive layer remain among the best-performing OPDs, with measured NEP values of 800 fW ($D^* \sim 10^{11} \text{ cm} \cdot \text{Hz}^{1/2} \text{ W}^{-1}$ at $B = 1 \text{ Hz}$) (8). In this work, we conducted detailed measurements on P3HT:ICBA OPDs using indium tin oxide (ITO)/polyethylenimine ethoxylated (PEIE) (14) and MoO_x/Ag electrodes (i.e., without injection-blocking layers) that achieve a level of performance in the visible spectral range that is statistically comparable to that of state-of-the-art low-noise SiPDs (Hamamatsu S1133, e.g., NEP $\sim 200 \text{ fW}$ and $D^* \sim 2 \times 10^{12} \text{ cm} \cdot \text{Hz}^{1/2} \text{ W}^{-1}$ at $B = 1.5 \text{ Hz}$), except response time, which is $\sim 35 \mu\text{s}$ and remains video-rate compatible.

Fig. 1. Comparison of SiPD and OPD performance.

(A) OPD geometry. (B) Measured spectral responsivity, EQE, external quantum efficiency. (C) Measured irradiance-dependent photocurrent and responsivity. LDR, linear dynamic range. (D) Measured rms noise current, NEP, and specific detectivity box plots (N represents the number of data points). Max, maximum; Min, minimum.



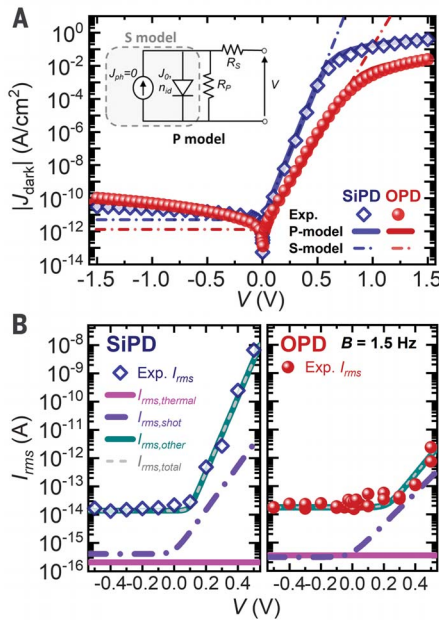


Fig. 2. Steady-state dark current density and electronic noise characteristics in SiPDs and OPDs. (A) Voltage-dependent dark current density. Exp., experimental. (B) Comparison of modeled and measured rms noise currents under reverse bias.

These large-area flexible OPDs can have superior design possibilities and performance to detect faint visible light, as demonstrated in photoplethysmography (PPG) experiments, a commonly used technique for wearable biometric monitoring.

A detailed methodology to characterize PDs (materials and methods) in the dark (fig. S1) and under illumination (fig. S2) enabled a statistically relevant comparison of the performance of S1133 SiPDs and OPDs (Fig. 1). S1133 SiPDs are comparable to S1133-01 SiPDs but have a color filter that yields similar spectral responsivity [$\mathcal{R}(\lambda)$] to that of OPDs (Fig. 1B).

First, the photocurrent, $I_{\text{ph}}(\phi_{\text{opt}})$, in SiPDs was measured at 525 nm to derive $\mathcal{R}(\phi_{\text{opt}}) \equiv I_{\text{ph}}(\phi_{\text{opt}})/\phi_{\text{opt}}$ and the NEP from signal-to-noise [$\text{SNR} = I_{\text{ph}}(\phi_{\text{opt}})/I_{\text{rms}}$] measurements (fig. S2). Figure S2A shows current transients at different ϕ_{opt} values with I_{rms} of 36 fA. Figure S2B shows that an interpolation of SNR values from these transients around unity yielded an NEP of 211 fW, different from an extrapolated NEP_{ex} of 131 fW, which was calculated by extrapolating $I_{\text{ph}}(\phi_{\text{opt}})$ to the I_{rms} value following the slope \mathcal{R} measured at higher ϕ_{opt} (as typically reported) (fig. S2C). This discrepancy arises because \mathcal{R} is no longer constant as $\phi_{\text{opt}} \rightarrow \text{NEP}$ (fig. S2D), an effect that may be related to photocurrent losses due to charge trapping (6). Multiple identical measurements yielded a wide distribution of $\mathcal{R}(\phi_{\text{opt}})$

values as $\phi_{\text{opt}} \rightarrow \text{NEP}$ (fig. S2E). Figure 1C shows the median and 25 to 75% percentiles of the $I_{\text{ph}}(\phi_{\text{opt}})$ and $\mathcal{R}(\phi_{\text{opt}})$ distributions, yielding $\mathcal{R}(\phi_{\text{opt}} > 1 \text{ pW}, 525 \text{ nm}) = 0.275 \text{ AW}^{-1}$, in agreement with data in Fig. 1B. However, for $\phi_{\text{opt}} < 1 \text{ pW}$, the width of the $\mathcal{R}(\phi_{\text{opt}})$ distribution increases, whereas its median decreases as $\phi_{\text{opt}} \rightarrow \text{NEP}$.

Noise measurements under the same experimental conditions on three SiPDs yielded a median I_{rms} of 24 fA (fig. S2F). Median I_{rms} and $\mathcal{R}(\phi_{\text{opt}} > 1 \text{ pW})$ values yielded a NEP_{ex} of 87 fW, leading to a large error when compared with the median NEP of 174 fW determined from multiple measurements (Fig. 1D, middle). A median D^* ($B = 1.5 \text{ Hz}$) value of $2.1 \times 10^{12} \text{ cm}^2 \text{ Hz}^{1/2} \text{ W}^{-1}$ is derived from NEP measurements (Fig. 1D, bottom). Figure S3 supports these measurements by showing that a 0.98-gigaohm resistor yields a median I_{rms} of 5.1 fA, in agreement with the thermal noise current ($I_{\text{thermal}} = \sqrt{4k_B T B R_p} = 5.0 \text{ fA}$ with $B = 1.5 \text{ Hz}$, where k_B is the Boltzmann constant and T is temperature).

Similar measurements in OPDs (Fig. 1C and fig. S2H) yielded larger variations of $\mathcal{R}(\phi_{\text{opt}})$ than those in SiPDs. Figure 1D (top) shows the distribution of I_{rms} values ($N = 104$ for six devices) (fig. S2I), with a median of 37 fA. The distribution of NEP values in OPDs has a median of 230 fW that is different from $\text{NEP}_{\text{ex}} = 136 \text{ fW}$ (Fig. 1D, middle). The median D^* in OPDs has a similar value ($1.8 \times 10^{12} \text{ cm}^2 \text{ Hz}^{1/2} \text{ W}^{-1}$) to that in S1133 SiPDs (Fig. 1D, bottom), thereby demonstrating a statistically comparable level of performance within the visible spectral range. Furthermore, fig. S4 shows that champion OPDs yielded an NEP as low as 63 fW and D^* as high as $8 \times 10^{12} \text{ cm}^2 \text{ Hz}^{1/2} \text{ W}^{-1}$ at 600 nm, similar to peak D^* values in S1133-01 SiPDs. Table S1 summarizes the performance parameters.

We now discuss the origins of these small I_{rms} and NEP values in OPDs. The low measurement bandwidth in these experiments suggests that the spectral noise characteristics are dominated by “pink” or $1/f$ noise, where f is frequency. The thermal noise and the shot noise that dominate at higher frequencies in the so-called white-noise region could not be measured directly because their values are below the noise floor of our equipment. However, their magnitude could be estimated from the values of the shunt resistance (R_p) and reverse saturation current density (J_0) of a PD. Measured $J_{\text{dark}}(V)$, where V is voltage, values in SiPDs and OPDs were fitted to the equivalent circuit model described by Prince (15, 16) (P-model in Fig. 2A). The P-model reduces to the well-known Shockley diode model (S-model, noted as SM): $J_{\text{SM, dark}}(V, T) = J_0 \left[\exp\left(\frac{qV}{n_{\text{id}} k_B T}\right) - 1 \right]$ (17) in the limit $R_p \rightarrow \infty$ (and $R_s \rightarrow 0$). In these

models, J_0 arises from charge-carrier generation and recombination in the dark, and n_{id} is the diode ideality factor, with a value determined by the type of charge-carrier recombination process that dominates in the dark (18, 19).

Figure 2A shows a fit to $J_{\text{dark}}(V)$, in SiPDs using values of $J_0 = 5 \text{ pA cm}^{-2}$, $n_{\text{id}} = 1.05$, and $R_p = 786 \text{ gigaohm}$. A large R_p in SiPDs allows $J_{\text{dark}}(V) \rightarrow J_{\text{SM, dark}}(V)$ as $V \rightarrow 0 \text{ V}$ (fig. S5); that is, $J_{\text{dark}}(V)$ approaches its thermodynamic limit but is never strictly equal to J_0 (fig. S6) because it arises from $J_{\text{SM, dark}}$ and a leakage current $V(R_p A_{\text{PD}})^{-1}$. Although a leakage current may arise from the combination of multiple phenomena, e.g., charge transport along grain boundaries or electronic defects across the photoactive layer, this model only includes those that depend linearly on V . Using J_0 and R_p values, a thermal noise power spectral density, $S_{\text{thermal}} = 4k_B T R_p^{-1}$, of $2.1 \times 10^{-32} \text{ A}^2 \text{ Hz}^{-1}$, and a shot-noise power spectral density at 0 V, $S_{\text{shot-diode}} = 4qJ_0 A_{\text{PD}}$ (20, 21), of $2.2 \times 10^{-31} \text{ A}^2 \text{ Hz}^{-1}$, were calculated, yielding a white-noise rms current, $I_{\text{rms, white-noise}} = (S_{\text{thermal}} + S_{\text{shot-diode}})^{1/2} B^{1/2}$, of 0.6 fA ($B = 1.5 \text{ Hz}$). As shown in Fig. 2B, white-noise contributions to I_{rms} are negligible compared with “other noise” ($I_{\text{rms, other}}$) contributions, presumably attributed to “pink” noise.

In OPDs, reproducible $J_{\text{dark}}(V)$ characteristics (fig. S7) enabled accurate estimation of R_p and J_0 values through detailed temperature- and optical power-dependent studies, analyzed by using the P-model (fig. S8). Figure 2A and figs. S6 and S8 show that quantitative agreement was obtained with values of $J_0 = 1.2 \text{ pA cm}^{-2}$, $n_{\text{id}} = 1.515$, and $R_p = 217 \text{ gigaohms}$ at $T = 297 \text{ K}$. Figure S8E shows that a charge-transfer state energy E_{CT} of $1.31 \pm 0.05 \text{ eV}$, derived from these experiments, is in agreement with values estimated (22) from the spectral characteristics of I_{ph} (fig. S8F) and with previously reported values (23). As in SiPDs, in these OPDs $J_{\text{dark}}(V) \rightarrow J_{\text{SM, dark}}(V)$, as $V \rightarrow 0 \text{ V}$ (Fig. S6). R_p and J_0 values yielded a S_{thermal} of $7.5 \times 10^{-32} \text{ A}^2 \text{ Hz}^{-1}$ and $S_{\text{shot-diode}}$ of

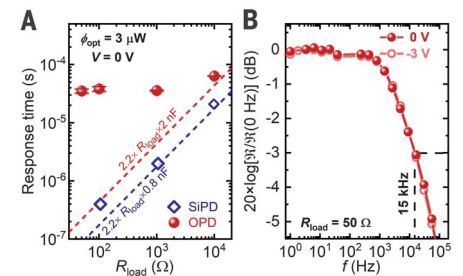


Fig. 3. Temporal response characteristics in SiPDs and OPDs. (A) Load resistance-dependent 10 to 90% rise and fall response times. (B) Frequency-dependent normalized responsivity at 525 nm.

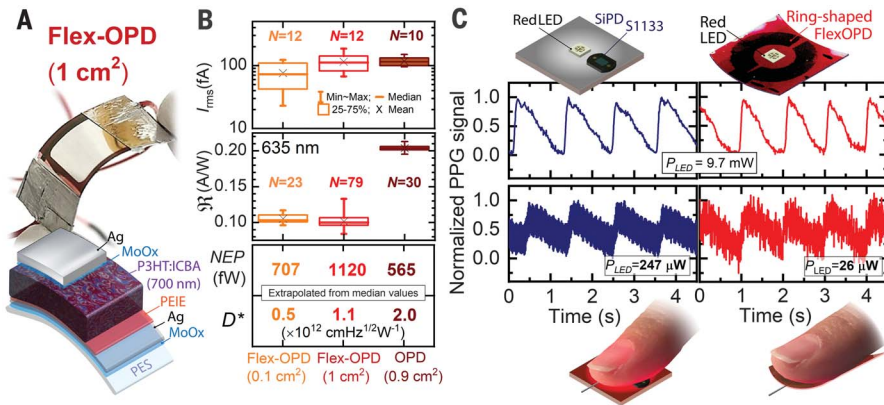


Fig. 4. Flex-OPDs and application in PPG. (A) Flex-OPD device geometry. PES, Polyethersulfone. (B) Box plots of rms noise current, responsivity, NEP, and specific detectivity in small- and large-area flex-OPDs and large-area OPDs. (C) Schematics of Si1133 SiPD and ring-shaped flex-OPD PPG array (top) and comparison of reflection-mode PPG signals measured on a finger (bottom) with a SiPD and a ring-shaped flex-OPD PPG array with the red LED operating at different electrical power values.

8.8×10^{-32} A² Hz⁻¹, leading to an $I_{rms,white-noise}$ of 0.5 fA ($B = 1.5$ Hz) and, therefore, to noise characteristics dominated by $I_{rms,other}$ (Fig. 2B). At higher-frequency bandwidths, $I_{rms,white-noise}$ values lead to an extrapolated white noise-limited NEP_{ex,wn} of ~2 fW and a $D^*_{ex,wn}$ of $(1.58 \times B^{1/2}) \times 10^{14}$ cm-Hz^{1/2} W⁻¹ for OPDs and, in agreement with reported values, NEP_{ex,wn} of 7 fW and $D^*_{ex,wn}$ of $(4 \times B^{1/2}) \times 10^{13}$ cm-Hz^{1/2} W⁻¹ for SiPDs.

Next, we investigated the photocurrent dynamics of OPDs in response to a light pulse by measuring the voltage drop across a load resistor (R_{load}). These measurements yielded a response time of 35 ± 10 μ s, independent of R_{load} for values smaller than 1 kilohm (Fig. 3A) and of applied electric field up to -32 V μ m⁻¹ (fig. S9). In this range, the resistance times capacitance (RC) time constant does not limit the OPD response time. As discussed in the supplementary materials, because the electric-field distribution in thick OPDs can be inhomogeneous (24), what transport mechanism limits the response time at this point is unclear. By contrast, and consistent with their data sheet, the response time of SiPDs is dominated by the RC time constant (Fig. 3A). Despite slower response times, OPDs still achieve a -3 -dB responsivity bandwidth of 15 kHz (Fig. 3B), making them video-rate compatible and well suited for biometric monitoring and other applications that require sub-kHz detection of faint optical signals. In such applications, OPDs will operate at frequencies dominated by $I_{rms,other}$. Whereas $I_{rms,white-noise}$ could be reduced by selecting materials with small J_0 values and by pursuing device optimization strategies to increase R_p , rational approaches to minimize $I_{rms,other}$ are less clear.

To investigate what factors contribute to $I_{rms,other}$, the influence of R_p was studied in P3HT:ICBA OPDs with thickness values in

the range of 200 to 700 nm. Figure S10 shows that OPDs with R_p values from 50 megaohm to 400 gigaohm yielded similar I_{rms} values, suggesting that $I_{rms,other}$ is independent of $I_{rms,thermal}$. Next, the influence of J_0 was studied on 960-nm-thick P3HT:PCBM OPDs (fig. S11) with $R_p = 6.5$ gigaohm and $J_0 = 35$ pA cm⁻². These values led to an $S_{thermal}$ of 3.0×10^{-30} A² Hz⁻¹ and $S_{shot-diode}$ of 2.7×10^{-30} A² Hz⁻¹ and yielded $I_{rms,white-noise} = 2.8$ fA. A median I_{rms} of 30.0 fA demonstrates that a 30 times increase in J_0 did not affect the magnitude of $I_{rms,other}$. However, measured NEP values (fig. S11D) in P3HT:PCBM OPDs are at least 10 times larger (in the picowatt range) than those in P3HT:ICBA OPDs, underscoring the importance of measuring $\mathfrak{R}(\phi_{opt} \rightarrow NEP)$.

Finally, 700-nm-thick P3HT:ICBA OPDs were studied with “inverted” and “conventional” geometries. Figure S12 shows that the use of Ca and poly(3,4-ethylenedioxythiophene) polystyrene sulfonate (PEDOT:PSS) in “conventional” OPDs yielded median I_{rms} of 227 fA and thus $I_{rms,other}$ of 226.9 fA, 4.5 times larger than in devices with ITO/PEIE and MoO_x/Ag electrodes. This result suggests that beyond affecting R_p values, contacts played an important role in minimizing $I_{rms,other}$ and highlight a challenge for the realization of OPDs with an extended spectral range and improved performance using state-of-the-art materials containing 1,1-dicyanomethylene-3-indanone end groups, which have been shown to chemically react with PEIE (25).

Finally, we illustrate how P3HT:ICBA OPDs present compelling advantages in their form factor over SiPDs. Large-area OPDs with custom shapes were fabricated on flexible substrates (flex-OPDs in Fig. 4A). The increased area, from 0.1 to ~1 cm², on rigid OPDs and flex-OPDs does not substantially increase either J_{dark} (V) (fig. S13) or I_{rms} (Fig. 4B, top). Using median I_{rms}

and \mathfrak{R} values, an extrapolated D^* value of 2.0×10^{12} cm-Hz^{1/2} W⁻¹ on 0.9-cm² OPDs on glass is in excellent agreement with the measured value in small-area devices.

Smaller \mathfrak{R} values on flex-OPDs compared with those on rigid OPDs resulted from series resistance, absorption, and reflection losses (~30%) caused by the semitransparent Ag/MoO_x electrode (fig. S13) and yielded D^* values between 2 and 4 times smaller in flex-OPDs when compared with values on OPDs on glass. Despite room for optimization, these values define state-of-the-art flexible OPDs (26, 27). Furthermore, an additional advantage of OPDs is the ease of creating photodetectors with complex shapes. To illustrate this, a ring-shaped flex-OPD, a geometry better suited to maximize the SNR and minimize the power consumption of PPG sensors (28), was fabricated with a total active area of 1 cm². PPG is a light-based technique that detects changes of blood vessel volume in the microvascular bed of tissues that can be used to assess cardiac and pulmonary function (29). Figure S14 demonstrates that a SiPD and an OPD of similar area yield comparable PPG signal quality. However, Fig. 4C demonstrates that a 1-cm² ring-shaped flex-OPD yields PPG signals with a SNR that is comparable to those measured with a 0.07-cm² SiPD even when the electrical power of the red light-emitting diode (LED) is reduced from 247 to 26 μ W. Therefore, large-area flex-OPDs could maximize functionality, reduce energy consumption, increase user comfort, and lower the cost of PPG sensors.

REFERENCES AND NOTES

- C. Perera, A. Zaslavsky, P. Christen, D. Georgakopoulos, *IEEE Comm. Surv. and Tutor.* **16**, 414–454 (2014).
- R. D. Jansen-van Vuuren, A. Armin, A. K. Pandey, P. L. Burn, P. Meredith, *Adv. Mater.* **28**, 4766–4802 (2016).
- G. Eppeldauer, J. E. Hardis, *Appl. Opt.* **30**, 3091–3099 (1991).
- F. P. García de Arquer, A. Armin, P. Meredith, E. H. Sargent, *Nat. Rev. Mater.* **2**, 16100 (2017).
- J. Lee et al., *ACS Energy Lett.* **4**, 1401–1409 (2019).
- Y. Fang, A. Armin, P. Meredith, J. Huang, *Nat. Photonics* **13**, 1–4 (2019).
- X. Gong et al., *Science* **325**, 1665–1667 (2009).
- S. Xiong et al., *ACS Appl. Mater. Interfaces* **9**, 9176–9183 (2017).
- A. Armin et al., *Laser Photonics Rev.* **8**, 924–932 (2014).
- A. Armin, R. D. Jansen-van Vuuren, N. Koidakis, P. L. Burn, P. Meredith, *Nat. Commun.* **6**, 6343 (2015).
- A. Pierre, I. Deckman, P. B. Lechêne, A. C. Arias, *Adv. Mater.* **27**, 6411–6417 (2015).
- P. E. Keivanidis, P. K. H. Ho, R. H. Friend, N. C. Greenham, *Adv. Funct. Mater.* **20**, 3895–3903 (2010).
- M. Kiehl, O. Dhez, G. Pecastaings, A. Curutchet, L. Hirsch, *Sci. Rep.* **6**, 39201 (2016).
- Y. Zhou et al., *Science* **336**, 327–332 (2012).
- W. J. Potscavage Jr., S. Yoo, B. Kippelen, *Appl. Phys. Lett.* **93**, 193308 (2008).
- M. B. Prince, *J. Appl. Phys.* **26**, 534–540 (1955).
- W. Shockley, *Bell Syst. Tech. J.* **28**, 435–489 (1949).
- T. Kirchartz, B. E. Pieters, J. Kirkpatrick, U. Rau, J. Nelson, *Phys. Rev. B* **83**, 115209 (2011).
- C. Göhler, A. Wagenpahl, C. Deibel, *Adv. Electron. Mater.* **4**, 1700505 (2018).
- A. van der Ziel, *Proc. IEEE* **58**, 1178–1206 (1970).
- R. L. Stratonovich, *Nonlinear Nonequilibrium Thermodynamics I: Linear and Nonlinear Fluctuation-Dissipation Theorems* (Springer, 1992).

22. K. Vandewal, K. Tvingstedt, A. Gadisa, O. Inganäs, J. V. Manca, *Phys. Rev. B* **81**, 125204 (2010).
23. L. Tzabari, J. Wang, Y.-J. Lee, J. W. P. Hsu, N. Tessler, *J. Phys. Chem. C* **118**, 27681–27689 (2014).
24. G. F. A. Dibb *et al.*, *Sci. Rep.* **3**, 3335 (2013).
25. L. Hu *et al.*, *J. Mater. Chem. A Mater. Energy Sustain.* **6**, 2273–2278 (2018).
26. A. Pierre, A. C. Arias, *Flex. Print. Electron.* **1**, 043001 (2016).
27. P. C. Y. Chow, T. Someya, *Adv. Mater.* **32**, 1902045 (2020).
28. H. Lee *et al.*, *Sci. Adv.* **4**, eaas9530 (2018).
29. J. Allen, *Physiol. Meas.* **28**, R1–R39 (2007).

ACKNOWLEDGMENTS

Funding: This research was supported in part by the Center for Organic Photonics and Electronics at Georgia Tech, by the Department of the Navy, Office of Naval Research (award nos. N00014-14-1-0580 and N00014-16-1-2520), through the MURI

Center for Advanced Organic Photovoltaics (CAOP), by the Air Force Office of Scientific Research (through award no. FA9550-16-1-0168), and by the Department of Energy National Nuclear Security Administration (NNSA) [award no. DE-NA0002576 through the Consortium for Nonproliferation Enabling Capabilities (CNEC) and award no. DE-NA0003921 through the Consortium for Enabling Technologies and Innovation (ETI)]. We also acknowledge the support of CONICYT (Chilean National Commission for Scientific and Technological Research) through the doctoral fellowship program “Becas Chile” (grant no. 72150387 for F.A.L.); support from COLCIENCIAS (Colombian Administrative Department of Science, Technology and Innovation) through the program Fulbright-Colciencias for V.A.R.-T.; support from the National Science Foundation through the Research Experiences for Undergraduates program for J.L.; and support from Brazil Scientific Mobility Program through an Academic Training Opportunities grant for L.D. **Author contributions:** C.F.-H. and B.K. conceptualized the work, acquired funding, managed the project, and supervised students; C.F.-H, W.-F.C., T.M.K., J.L., L.D.,

V.A.R.-T., and F.A.L. conducted investigations, verification, and data curation. C.F.-H. developed methodology, conducted data analysis, developed software to acquire and analyze data, and wrote the original draft. All authors contributed in reviewing and editing the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6517/698/suppl/DC1

Materials and Methods

Supplementary Text

Figs. S1 to S14

Table S1

Reference (30)

18 November 2019; resubmitted 23 July 2020

Accepted 11 September 2020

10.1126/science.aba2624

TOPOLOGICAL OPTICS

Nonlinearity-induced photonic topological insulator

Lukas J. Maczewsky¹, Matthias Heinrich¹, Mark Kremer¹, Sergey K. Ivanov^{2,3}, Max Ehrhardt¹, Franklin Martinez¹, Yaroslav V. Kartashov^{3,4}, Vladimir V. Konotop^{5,6}, Lluís Torner^{4,7}, Dieter Bauer¹, Alexander Szameit^{1,*}

A hallmark feature of topological insulators is robust edge transport that is impervious to scattering at defects and lattice disorder. We demonstrate a topological system, using a photonic platform, in which the existence of the topological phase is brought about by optical nonlinearity. The lattice structure remains topologically trivial in the linear regime, but as the optical power is increased above a certain power threshold, the system is driven into the topologically nontrivial regime. This transition is marked by the transient emergence of a protected unidirectional transport channel along the edge of the structure. Our work studies topological properties of matter in the nonlinear regime, providing a possible route for the development of compact devices that harness topological features in an on-demand fashion.

With the discovery of topological insulators (TIs), materials science ushered in a new era of physics (1, 2). Solid-state TIs prohibit electrons from traversing the bulk while simultaneously supporting chiral surface currents that are topologically protected from scattering at defects and disorder (3). Soon after their first realizations in condensed matter systems (4, 5), topological concepts were implemented across other fields of physics, resulting in the experimental demonstrations of topological dynamics in various platforms (6–10).

Ongoing efforts have been directed toward connecting topology and nonlinearity (11), which may enable topological Mott insulators (12), interaction-induced TIs such as the Kondo insulator (13), and non-Abelian TIs (14) and may even drive the formation of topological solitons (15, 16). Only recently, bulk soliton formation within a topological bandgap was demonstrated experimentally (17). However, work on nonlinear TIs and their corresponding robust edge transport has thus far had to cope with various limitations of the respective implementations that only allowed for experiments with purely linear edge state dynamics (17–19).

We theoretically and experimentally demonstrate a nonlinearity-induced photonic TI, showing how nonlinearity can drive an initially topologically trivial system into a tran-

sient topological phase, where protected chiral edge states exist (Fig. 1A). In contrast to soliton formation in an arrangement with preexisting topological features (17), it is the action of nonlinearity itself that establishes nontrivial topology. We use lattices of coupled optical waveguides as a platform to explore nonlinear physics (20–22). In particular, we make use of a modified anomalous Floquet TI arrangement (23–25), which can exhibit unidirectional edge transport for certain parameters. Notably, the Chern number \mathcal{C} of this structure generally remains zero. Instead, the topological phase of this system is adequately described by a winding number \mathcal{W} that counts the number of topologically protected edge states (25).

In the tight-binding regime, the light dynamics in our finite $(2 + 1)$ dimensional system is modeled by the discretized Schrödinger equation (26)

$$i \frac{\partial}{\partial z} a_n(z) + \sum_{\langle m \rangle} H_{m,n}(z) a_m(z) + \gamma |a_n(z)|^2 a_n(z) = 0 \quad (1)$$

for the field amplitudes $a_n(z)$, where $H_{m,n}(z)$ is the linear tight-binding Hamiltonian, describing the z -dependent coupling from site n to a nearest neighbor m (indicated by $\langle m \rangle$ in the sum) in the two-dimensional lattice. In turn, the quantity γ describes the strength of our Kerr-type nonlinearity. Evidently, Eq. 1 is similar to the Gross-Pitaevskii equation, with the temporal evolution being replaced by spatial dynamics in z along the waveguides, and $\gamma > 0$ describing focusing or positive nonlinearity (26). We consider a $(2 + 1)$ -dimensional spatially discrete system, which does not exhibit chiral edge states and remains topologically trivial for low-intensity beams. However, at sufficient excitation power, the term $\gamma |a_n(z)|^2 a_n(z)$ becomes important in Eq. 1. As we will show in our experiments, this can

drive the system into a topologically nontrivial phase and promote the formation of nonlinear chiral edge states.

The underlying principle can be understood by considering two interacting waveguides that form a directional coupler (Fig. 1B). Whereas such a coupler composed of identical waveguides exhibits the characteristic sinusoidal intensity oscillation and the associated periodic full-power transfer between the waveguides, introducing a detuning between their effective refractive indices forces a certain fraction of light to remain in the initially excited guide at all times. In the linear regime, the maximum reached in the second waveguide therefore remains well below the total input power. However, when launching high-power light into the lower-index guide, a positive nonlinearity ($\gamma > 0$) can yield an intermittent phase-matching with the neighboring higher-index guide. Consequently, the fraction of power that can be transferred increases with the injected power (Fig. 1C). Such photonic waveguide structures, fabricated by means of femtosecond laser direct inscription [(27); see methods (28)], serve as building blocks for our structure and indeed show this behavior when excited with intense laser pulses (Fig. 1C). For low input powers ($P_{\text{in}} \approx 50$ kW peak power), only about 44% of the light is transferred. We were able to increase this transfer efficiency up to 73% in the nonlinear regime at a power of $P_{\text{NL}} = 2.9$ MW. The simulated transfer dynamics illustrates how such an increased transfer ratio is achieved despite the focusing nature of the nonlinearity by judiciously tailoring the length of the interaction region (Fig. 1, D and E). This nonlinear switching is the basis for our implementation of a nonlinear photonic TI (see section 1 of (28) for details).

Our topological structure is composed of optical waveguides arranged in a bipartite square lattice (Fig. 2A). The individual channels are selectively brought into evanescent contact with one another so as to form a cyclic discrete coupling pattern. At any distinct propagation step j , represented by a dedicated directional coupler, light from each waveguide is partially transferred with ratio t to only one specific nearest neighbor. Additionally, the on-site potential (i.e., the effective refractive index) is modulated for each successive step (Fig. 2B). This Floquet driving protocol establishes periodicity in the transverse (x, y) as well as the longitudinal (z) direction, with (driving) periods d and Z , respectively. Notably, the system enters a nontrivial topological phase when the transfer ratio in each step exceeds $t > 50\%$ (23). Above this topological phase transition, the winding number assumes the value of $\mathcal{W} = 1$ (25) [see methods (28) for a detailed discussion].

Although the transfer ratio of each driving step determines the topological phase of the

¹Institut für Physik, Universität Rostock, Albert-Einstein-Str. 23, 18059 Rostock, Germany. ²Moscow Institute of Physics and Technology, Institutsky lane 9, Dolgoprudny, Moscow Region 141700, Russia. ³Institute of Spectroscopy, Russian Academy of Sciences, Fizicheskaya Str., 5, Troitsk, Moscow 108840, Russia. ⁴ICFO-Institut de Ciències Fotoniques, The Barcelona Institute of Science and Technology, 08860 Castelldefels (Barcelona), Spain. ⁵Departamento de Física, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, Lisboa 1749-016, Portugal. ⁶Centro de Física Teórica e Computacional, Faculdade de Ciências, Universidade de Lisboa, Campo Grande, Lisboa 1749-016, Portugal. ⁷Universitat Politècnica de Catalunya, 08034 Barcelona, Spain. *Corresponding author. Email: alexander.szameit@uni-rostock.de

system, the band structure also depends on the microdynamics of the electric field. We therefore analyze the linear dynamics of all four individual steps j of length $Z/4$ by considering the (k_x, k_y) -space of the time-discretized linear bulk Hamiltonian. Following a Bloch ansatz to the real-space Hamiltonian

of Eq. 1 of the infinitely extended bipartite lattice, we find

$$H_B(\mathbf{k}, z) = \sum_{j=1}^4 \begin{bmatrix} \delta^{(1)}(z) & c_j(z)e^{i\mathbf{b}_j \cdot \mathbf{k}} \\ c_j(z)e^{-i\mathbf{b}_j \cdot \mathbf{k}} & \delta^{(2)}(z) \end{bmatrix} \quad (2)$$

where the vectors $\{\mathbf{b}_j\}$ are defined as $\mathbf{b}_1 = -\mathbf{b}_3 = (d/2, 0)$ and $\mathbf{b}_2 = -\mathbf{b}_4 = (0, d/2)$, with the transverse lattice constant d (Fig. 2C) and $\mathbf{k} = (k_x, k_y)$. During the j th step of the period, the coupling $c_j = c$, whereas the other three are set to zero. The detuning $\delta^{(1)}$ of one sublattice is set to value δ in steps 1 and 4 and remains zero for steps 2 and 3, whereas its counterpart $\delta^{(2)}$ is fixed to δ in steps 2 and 3 and set to zero for steps 1 and 4 instead [see Fig. 2, B and C, and section 2 of (28)]. We start our analysis by choosing the coupling coefficient as $c = 1.5 \frac{\pi}{Z}$ and a detuning value of $\delta = 3.2 \frac{\pi}{Z}$, which correspond to the parameters used for the detuned waveguide pair from Fig. 1B. In this case, the transfer ratio t in each step is only 44%, placing the system well within the topologically trivial phase ($\mathcal{C} = \mathcal{W} = 0$). The corresponding quasi-energy spectrum $\varepsilon(\mathbf{k})$ of the system exhibits dispersive bands separated by a trivial bandgap (Fig. 2D), and the edge quasi-energy spectrum $\varepsilon(k_x)$ of the truncated array is devoid of chiral edge states.

The situation changes when nonlinearity comes into play. The dynamics between linear coupling c and nonlinear self-focusing can be described by an effective nonlinear detuning value $\delta_{\text{NL}}^{\text{eff}} = \delta_{\text{lin}} + \delta_{\text{NL}}(|a_n|^2)$ (26), where δ_{lin} is the permanent detuning of the sites introduced during the fabrication process ($\delta^{(1)}$ and $\delta^{(2)}$). In the preliminary experiments (Fig. 1C), we find $\delta_{\text{NL}}^{\text{eff}} = 1.2 \frac{\pi}{Z}$. The electronic Kerr response of the host material fused silica guarantees that any such propagation effects are virtually instantaneous and strictly local. Notably, the Floquet theorem is, in general, not applicable to nonlinear differential equations such as Eq. 1 because their dynamics feed back into the nonlinear part of the Hamiltonian and therefore may deviate from the time periodicity of the linear part (29). However, a judicious choice of interaction distances within the couplers allows us to ensure that the nonlinear dynamics remain indeed periodic in the evolution (i.e., z in our case), as does the entire nonlinear Hamiltonian. It is therefore possible to calculate a meaningful quasi-energy spectrum of the Floquet lattice with $\delta_{\text{NL}}^{\text{eff}}$ (30) (Fig. 2E). When the nonlinear effective detuning increases the transfer ratio above 50%, a nontrivial topological phase is locally established. Although the band retains its essential characteristics, a chiral edge state emerges (marked magenta in Fig. 2E). For this effective nonlinear coupling, the local winding number can be computed to be $\mathcal{W} = 1$ (23–25). This effective spectrum is only present as long as the power is sufficient to induce the required effective detuning. As such, the resulting chiral edge states exist exclusively because of nonlinear light dynamics.

To observe the formation of these nonlinear chiral edge states, we fabricated extended samples composed of 36 waveguides, with the four partial coupling steps of one and two full

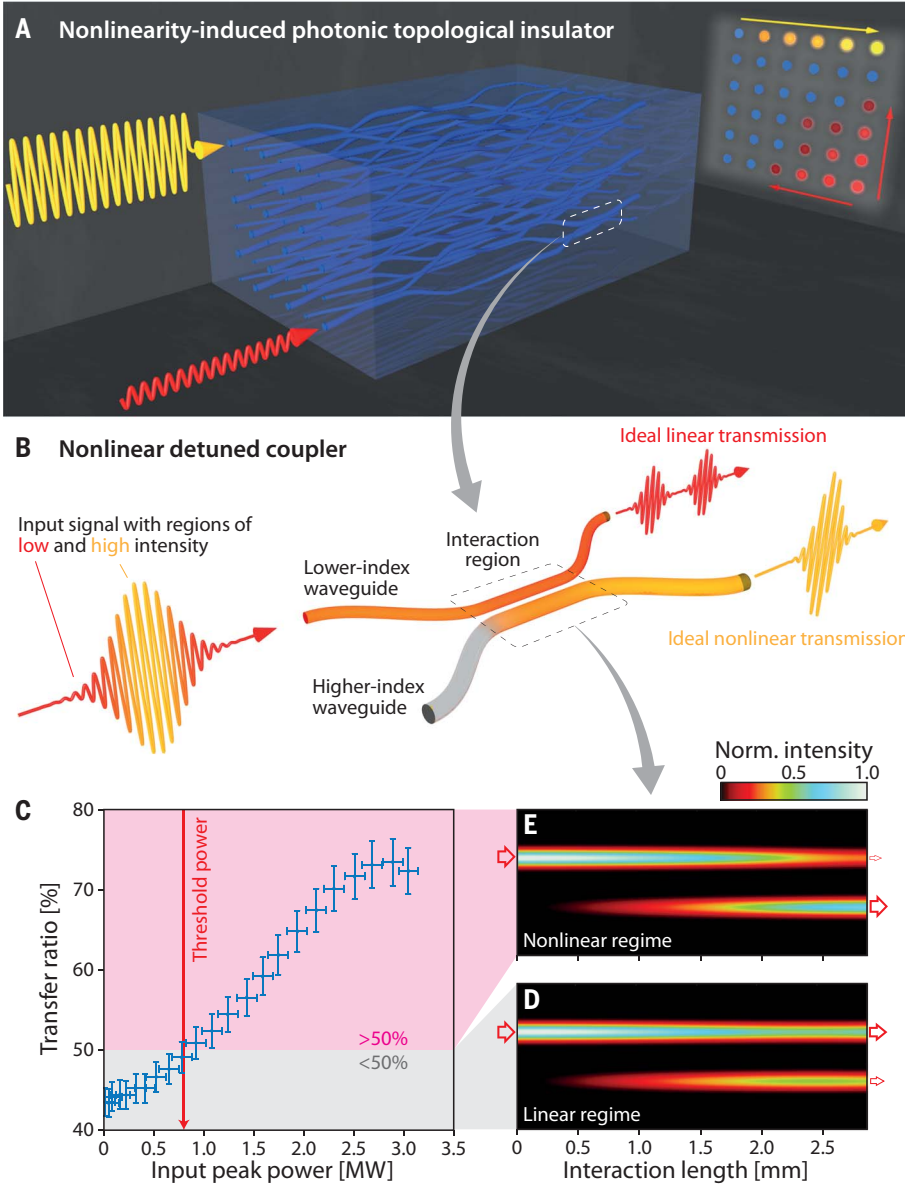
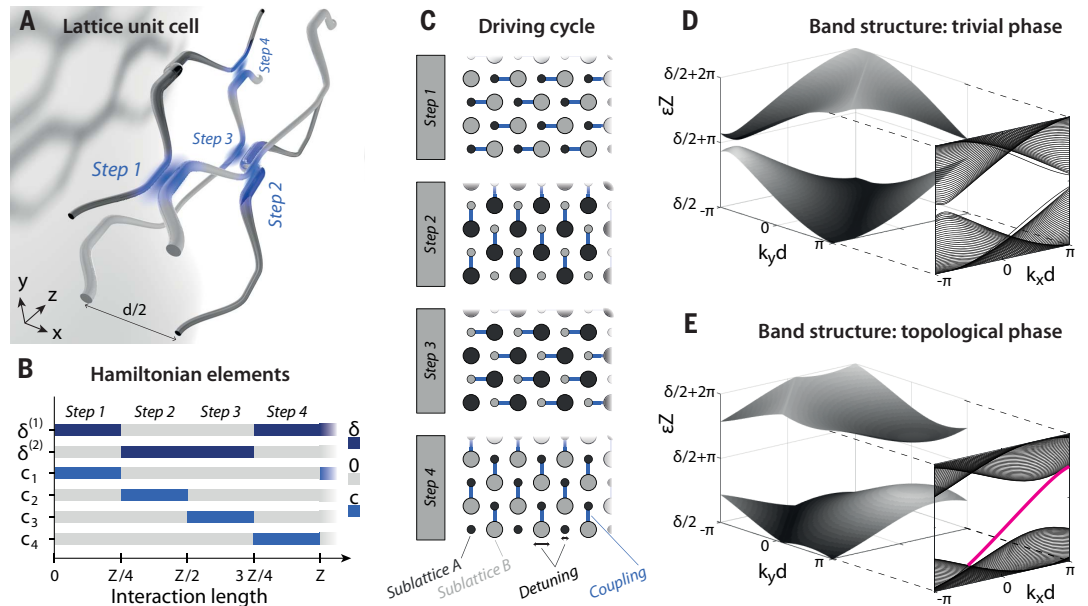


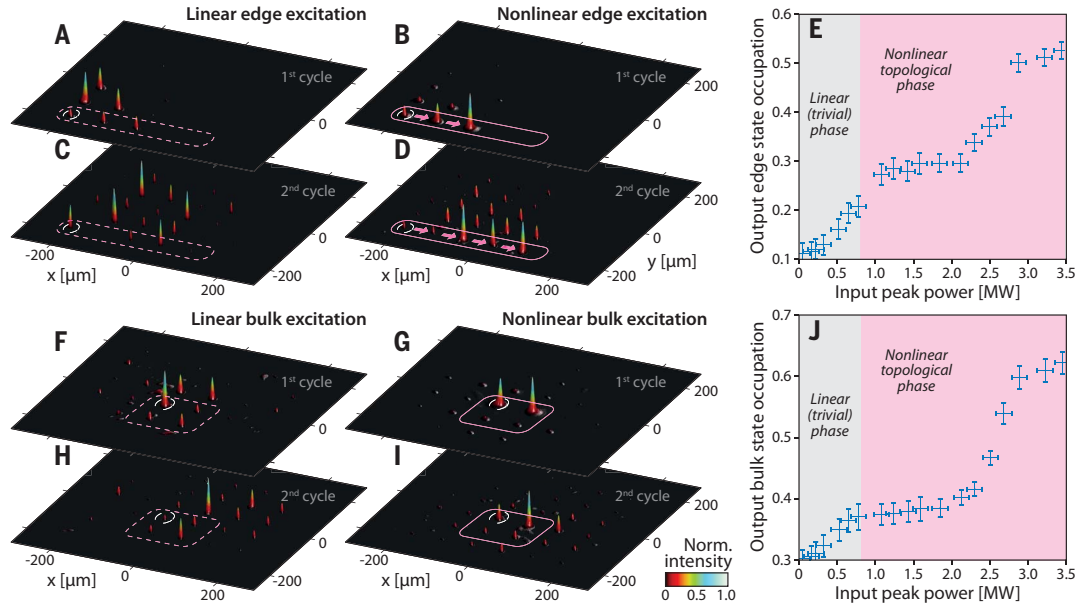
Fig. 1. Topology through nonlinearity. (A) Nonlinearity-induced photonic TI. Low-power edge excitations (red) experience diffraction, indicating a topologically trivial regime. By contrast, high-power light (yellow) produces a self-guided unidirectional edge state that travels along the perimeter of the structure. (B) Nonlinear directional detuned coupler. An intense laser pulse is launched into the lower-refractive index waveguide (smaller diameter) of a detuned coupler. A focusing Kerr nonlinearity allows the high-intensity part of the pulse to momentarily compensate for the detuning, whereby light is transferred to the higher-refractive index waveguide (larger diameter). By contrast, the lower-intensity parts exhibit linear dynamics and remain in the lower-index waveguide. (C) Measured power-dependent transfer ratio in a detuned coupler. As a guide to the eye, the regions of below- and above-50% transfer are shaded gray and magenta, respectively. Transfer ratios above 50% enable the topologically nontrivial phase of our lattice. Error bars indicate the measurement uncertainty of the individual data points. (D and E) Simulated light propagation in a detuned coupler for (D) low and (E) high intensities. The arrows schematically indicate the injected intensity (left) as well as the relative amounts of light emerging from the two output channels of the coupler (right).

Fig. 2. Driving protocol.

(A) Waveguide-based implementation of the coupling sequence. The detuning is indicated by the respective waveguide diameters (lower and higher refractive index for smaller and larger diameter, respectively). (B) Visualization of the parameters of Eq. 2. The higher-index guides are characterized by a detuning δ , and selective interaction between adjacent guides is indicated by the coupling coefficient c . (C) Schematic coupling sequence. The two species of the bipartite lattice (gray and black) are selectively allowed to interact, as indicated by the blue lines. The on-site potential of each waveguide (symbolized by its diameter, with larger diameter indicating greater potential) is modulated along the propagation direction. (D and E) Three-dimensional bulk band structure of the driven lattice is periodic in the quasi-momenta k_x , k_y and the quasi-energy ϵ , exhibiting a pronounced bandgap. In the presence of an edge, the band structure features (D) a trivial bandgap in the linear regime and (E) a chiral edge state (solid magenta line) connecting the bands, when nonlinearity compensates the detuning of the on-site potential and drives the system into its topological phase.

**Fig. 3. Experimental observation of nonlinearity-induced topological dynamics.**

(A to D) Intensity distribution for single-site edge excitation for [(A) and (B)] one and [(C) and (D)] two driving periods. Bulk-diffractive behavior of a low-power edge excitation is shown in (A) and (C). Unidirectional edge transport at high power (3.2 MW), as indicated by magenta arrows, is shown in (B) and (D). (E) Power-dependent edge occupation after two driving periods. (F to J) Intensity distribution for single-site bulk excitation for [(F) and (G)] one driving period and [(H) and (I)] two driving periods. Bulk dynamics around the excitation site for low-power excitation are shown in (F) and (H). Suppression of bulk diffraction above the nonlinear phase transition is shown in (G) and (I), confirmed by (J) a marked increase of the bulk state occupation. In (A) to (D) and (F) to (I), the white dashed circles indicate the excited waveguides. In (E) and (J), error bars indicate the measurement uncertainty of the individual data points.



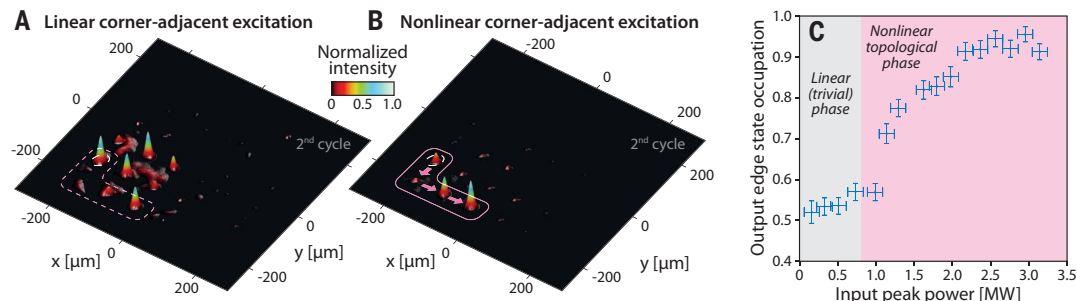
Floquet periods fabricated in samples up to 15 cm in length. The required linear detuning sequence was implemented by choosing appropriate writing speeds [see methods (28)]. In each coupling step, the interaction region was tailored such that an identical effective linear coupling was achieved [for details, see section 1 of (28)]. When exciting an edge waveguide with low-power light at $P_{\text{lin}} =$

50 kW (i.e., $t < 50\%$), diffraction into the bulk of the lattice and especially in all surrounding edge sites occurs, indicating the topologically trivial phase (see Fig. 3, A and C). By contrast, as the input peak power is increased up to $P_{\text{NL}} = 3.5$ MW (i.e., $t > 50\%$), bulk diffraction is suppressed and the chiral edge state forms (Fig. 3, B and D). The observed edge state occupation for two driving periods, determined

as the ratio $I_{\text{edge}}/I_{\text{total}}$ of the intensity of the five involved edge lattice sites (indicated by the magenta outline in Fig. 3D) over the total intensity in the lattice, clearly demonstrates the transition into the topological phase at the threshold power of 0.8 MW (Fig. 3E; see also section 4 of (28) and fig. S8). By contrast, a bulk excitation (Fig. 3, D to J) contracts toward its initially excited site at higher launched

Fig. 4. Evolution of a nonlinearity-induced topologically protected edge state around a corner of our system. (A and B) Experimentally observed intensity patterns after two full driving cycles for (A) linear and (B) nonlinear evolution, respectively. The excited waveguide is marked with a white dashed circle. **(C)** Power-

dependent experimental edge state occupation for the corner-adjacent excitation. Error bars indicate the measurement uncertainty of the individual data points.



powers (Fig. 3, G and I), resulting in a marked increase of the bulk state occupation, calculated as the ratio $I_{\text{bulk}}/I_{\text{total}}$ of the intensity at the linear bulk state lattice sites outlined by a rounded square in Fig. 3I over the total intensity. Crucially, this contraction does not arise from a nonlinear cancellation of coupling (17), as would occur in the case of soliton formation, but is instead a signature of bulk states that arise in the nonlinearity-induced topological phase of the lattice and localize light similar to the behavior known from conventional linear TIs.

Similar to their counterparts in conventional linear TIs, our nonlinear edge states are topologically protected against scattering, as evidenced by their scatter-free unidirectional propagation around a corner of the lattice (see Fig. 4). Despite the transient nature of the underlying topological phase, the edge light transport is robust and could persist even in the presence of imposed coupling disorder and artificial defects—as long as the power of the propagating beam is sufficient to induce the required amount of effective coupling. Therefore, nonlinearly induced topologically protected edge state transport can also be terminated at will by introducing losses so as to reduce the intensity below the phase transition threshold. Alternatively, this reverse transition can also be brought about at a predetermined propagation distance by making use of the gradual decrease of peak intensity owing to propagation losses or dispersive pulse broadening.

Notably, the ability to terminate topological protection at will, and at a location encoded by the initially injected power or pulse prechirp, addresses one of the main challenges hindering the application of topological systems, namely, how to efficiently extract signals at the destination of their topologically protected journey.

REFERENCES AND NOTES

1. C. L. Kane, E. J. Mele, *Phys. Rev. Lett.* **95**, 146802 (2005).
2. B. A. Bernevig, S. C. Zhang, *Phys. Rev. Lett.* **96**, 106802 (2006).
3. M. Z. Hasan, C. L. Kane, *Rev. Mod. Phys.* **82**, 3045–3067 (2010).
4. M. König et al., *Science* **318**, 766–770 (2007).
5. D. Hsieh et al., *Nature* **452**, 970–974 (2008).
6. Z. Wang, Y. Chong, J. D. Joannopoulos, M. Soljačić, *Nature* **461**, 772–775 (2009).
7. M. C. Rechtsman et al., *Nature* **496**, 196–200 (2013).
8. M. Hafezi, E. A. Demler, M. D. Lukin, J. M. Taylor, *Nat. Phys.* **7**, 907–912 (2011).
9. L. Lu, J. D. Joannopoulos, M. Soljačić, *Nat. Photonics* **8**, 821–829 (2014).
10. T. Ozawa et al., *Rev. Mod. Phys.* **91**, 015006 (2019).
11. S. Rachel, *Rep. Prog. Phys.* **81**, 116501 (2018).
12. S. Raghu, X. L. Qi, C. Honerkamp, S. C. Zhang, *Phys. Rev. Lett.* **100**, 156401 (2008).
13. V. Alexandrov, M. Dzero, P. Coleman, *Phys. Rev. Lett.* **111**, 226403 (2013).
14. J. C. Y. Teo, C. L. Kane, *Phys. Rev. B* **89**, 085101 (2014).
15. Y. Lumer, Y. Plotnik, M. C. Rechtsman, M. Segev, *Phys. Rev. Lett.* **111**, 243905 (2013).
16. D. Leykam, Y. D. Chong, *Phys. Rev. Lett.* **117**, 143901 (2016).
17. S. Mukherjee, M. C. Rechtsman, *Science* **368**, 856–859 (2020).
18. Y. Hadad, J. C. Soric, A. B. Khanikaev, A. Alù, *Nat. Electron.* **1**, 178–182 (2018).
19. F. Zangeneh-Nejad, R. Fleury, *Phys. Rev. Lett.* **123**, 053902 (2019).
20. H. S. Eisenberg, Y. Silberberg, R. Morandotti, A. R. Boyd, J. S. Aitchison, *Phys. Rev. Lett.* **81**, 3383–3386 (1998).
21. J. W. Fleischer, M. Segev, N. K. Efremidis, D. N. Christodoulides, *Nature* **422**, 147–150 (2003).
22. A. Szameit et al., *Opt. Express* **13**, 10552–10557 (2005).
23. L. J. Maczewsky, J. M. Zeuner, S. Nolte, A. Szameit, *Nat. Commun.* **8**, 13756 (2017).

24. S. Mukherjee et al., *Nat. Commun.* **8**, 13918 (2017).
25. M. S. Rudner, N. H. Lindner, E. Berg, M. Levin, *Phys. Rev. X* **3**, 031005 (2014).
26. D. N. Christodoulides, R. I. Joseph, *Opt. Lett.* **13**, 794–796 (1988).
27. A. Szameit, S. Nolte, *J. Phys. B* **43**, 163001 (2010).
28. See supplementary materials.
29. V. Kapoor, M. Ruggenthaler, D. Bauer, *Phys. Rev. A* **87**, 042521 (2013).
30. Y. Hadad, A. B. Khanikaev, A. Alù, *Phys. Rev. B* **93**, 155112 (2016).

ACKNOWLEDGMENTS

We thank C. Otto for preparing the high-quality fused silica samples used in this work. **Funding:** A.S. acknowledges funding from the Deutsche Forschungsgemeinschaft (grants SCH 612/6-1, SZ 276/12-1, BL 574/13-1, SZ 276/15-1, and SZ 276/20-1) and the Alfred Krupp von Bohlen and Halbach Foundation. V.V.K. acknowledges support from the Portuguese Foundation for Science and Technology (FCT) under contract no. UIDB/00618/2020. Y.V.K. and L.T. acknowledge support from the Government of Spain (Severo Ochoa CEX2019-000910-S), Fundació Cellex, Fundació Mir-Puig, and Generalitat de Catalunya (CERCA). Y.V.K. and S.K.I. acknowledge funding of this study by RFBR and DFG according to research project no. 18-502-12080. **Author contributions:** L.J.M., M.H., and A.S. developed the driving protocol and its experimental implementation. L.J.M. carried out the experiments. M.E. and F.M. provided technical support and assisted in the sample fabrication. Y.V.K., S.K.I., and L.T. formulated the numerical model and conducted the continuous simulations. D.B., M.K., L.J.M., and V.V.K. formulated the linearized Hamiltonian model. A.S. supervised the project. All authors discussed the results and co-wrote the paper. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All experimental data and simulations not provided in the text or the supplementary materials can be found at the Rostock University Publication Server repository (31).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6517/701/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S8
References (31–35)

7 June 2020; accepted 17 September 2020
10.1126/science.abd2033

CLIMATE CHANGE

Global food system emissions could preclude achieving the 1.5° and 2°C climate change targets

Michael A. Clark^{1*}, Nina G. G. Domingo², Kimberly Colgan², Sumil K. Thakrar², David Tilman^{3,4}, John Lynch⁵, Inês L. Azevedo^{6,7}, Jason D. Hill²

The Paris Agreement's goal of limiting the increase in global temperature to 1.5° or 2°C above preindustrial levels requires rapid reductions in greenhouse gas emissions. Although reducing emissions from fossil fuels is essential for meeting this goal, other sources of emissions may also preclude its attainment. We show that even if fossil fuel emissions were immediately halted, current trends in global food systems would prevent the achievement of the 1.5°C target and, by the end of the century, threaten the achievement of the 2°C target. Meeting the 1.5°C target requires rapid and ambitious changes to food systems as well as to all nonfood sectors. The 2°C target could be achieved with less-ambitious changes to food systems, but only if fossil fuel and other nonfood emissions are eliminated soon.

The goal of the Paris Agreement is to limit average global temperature increases above preindustrial levels to “well below 2°C” and to pursue efforts to “limit increase to 1.5°C.” Achieving either goal requires large and rapid reductions in greenhouse gas (GHG) emissions (1). To date, most efforts have focused on reducing GHG emissions from fossil fuel combustion in electricity production, transportation, and industry. Renewable energy sources, electric vehicles, improved efficiency, and other innovations and behavioral changes could eliminate most of these emissions, and carbon capture and sequestration could reduce atmospheric levels of previously emitted carbon. However, eliminating all emissions from these sectors may not be sufficient to meet the 1.5° and 2°C temperature targets. The global food system is also a major source of GHG emissions, emitting ~30% of the global total (2, 3). Nevertheless, reducing food-related emissions has received less attention, perhaps because these emissions might seem to be an unavoidable environmental cost of feeding humanity.

The global food system generates GHG emissions from multiple sources. Major sources include land clearing and deforestation, which release carbon dioxide (CO₂) and nitrous oxide (N₂O); production and use of fertilizers and other agrichemicals, which emit CO₂, N₂O, and methane (CH₄); enteric fermentation during the production of ruminants (cows, sheep, and

goats), which emits CH₄; production of rice in paddies, which emits CH₄; livestock manure, which emits N₂O and CH₄; and combustion of fossil fuels in food production and supply chains, which emits CO₂. In total, global food system emissions averaged ~16 billion tonnes (Gt) CO₂ equivalents year⁻¹ from 2012 to 2017 (4).

Here, we forecast GHG emissions from the global food system and assess whether they are compatible with the 1.5° and 2°C targets. We forecast emissions as a function of per capita diets (what is eaten and how much), the GHG intensity of various types of foods (emissions per unit of food produced, as estimated through life cycle assessment), and global population size. We assume that food systems continue to transition along trajectories of the past 50 years, which we refer to as business-as-usual (5, 6). This business-as-usual forecast makes straightforward assumptions: (i) per capita dietary composition and caloric consumption continue to change as countries become more affluent (5); (ii) crop yields, which influence how much land is converted to cropland, increase along recent trajectories (5); (iii) global population increases along the United Nation's medium-fertility pathway (7); and (iv) the GHG intensity of foods (8) and the rates of food loss and waste (9) remain constant through time.

GHG emissions from the global food system largely occur from food production and from land being cleared for food production. Emissions from food production are calculated by pairing life cycle assessment estimates of the GHG emissions per unit of each type of food (8) with their forecasted total global demand, and these estimates include emissions from activities such as production of agricultural inputs, fertilizer application, and animal husbandry. Our estimates of emissions from supply chains do not include emissions from transportation, processing, packaging, retail, and preparation, which in total account for a

minor fraction (~17%) of total food system emissions (10). Emissions from clearing land for food production are estimated by projecting crop yields, combining these with dietary projections to calculate annual rates of agricultural land-cover change, and pairing annual rates of agricultural land-cover change with Intergovernmental Panel on Climate Change (IPCC) Tier 1 estimates of GHG emissions from land clearing or carbon storage in biomass and soil after land abandonment (11, 12).

We next determine the maximum allowable cumulative GHG emissions from all human activities from 2020 onward that are compatible with having a 67 or 50% chance of meeting the 1.5° and 2°C targets, on the basis of the thresholds set in the IPCC Special Report on Global Warming of 1.5°C (13). We call these the emissions limits. To accurately incorporate CH₄ into the cumulative emissions framework, we report emissions as global warming potential (GWP*) CO₂ warming-equivalents (CO₂-we) (14). We also show results with the more commonly used GWP100 (100-year GWP) metric in data S2. To have a 67% chance of meeting the 1.5° and 2°C targets, the cumulative emissions limits are 500 and 1405 Gt CO₂-we, respectively. For a 50% chance of meeting the targets, the emissions limits are 705 and 1816 Gt CO₂-we, respectively (see supplementary materials).

Our analysis suggests that reducing GHG emissions from the global food system will likely be essential to meeting the 1.5° or 2°C target. Our estimate of cumulative business-as-usual food system emissions from 2020 to 2100 is 1356 Gt CO₂-we (Fig. 1). As such, even if all non-food system GHG emissions were immediately stopped and were net zero from 2020 to 2100, emissions from the food system alone would likely exceed the 1.5°C emissions limit between 2051 and 2063 (date range reflects uncertainties in the 1.5°C emissions limit; see supplementary materials). Further, given our estimate of food system emissions, maintaining a 67% chance of meeting the 2°C target would require keeping cumulative nonfood emissions to <50 Gt CO₂-we in total over the next 80 years. This is slightly more than 1 year of current GHG emissions from non-food system activities (4). Maintaining a 50% chance of meeting the 2°C target would allow for 455 Gt CO₂-we in total from nonfood emissions, which is 9 years of current nonfood emissions (4). These general trends hold even if emissions from fossil fuel use in the global food system were also to be immediately halted (see supplementary materials).

We next explore how global food system GHG emissions might be reduced through five strategies that target food supply and demand: (i) globally adopting a plant-rich diet [here modeled as a diet rich in plant-based foods that contains moderate amounts of dairy, eggs,

¹Oxford Martin School and Nuffield Department of Population Health, University of Oxford, Oxford, UK. ²Department of Bioproducts and Biosystems Engineering, University of Minnesota, St. Paul, MN, USA. ³Department of Ecology, Evolution, and Behavior, University of Minnesota, St. Paul, MN, USA. ⁴Bren School of Environmental Science and Management, University of California, Santa Barbara, CA, USA. ⁵Department of Physics, University of Oxford, Oxford, UK. ⁶Department of Energy Resources Engineering, Stanford University, Stanford, CA, USA. ⁷Woods Institute for the Environment, Stanford, CA, USA.

*Corresponding author. Email: michael.clark@ndph.ox.ac.uk

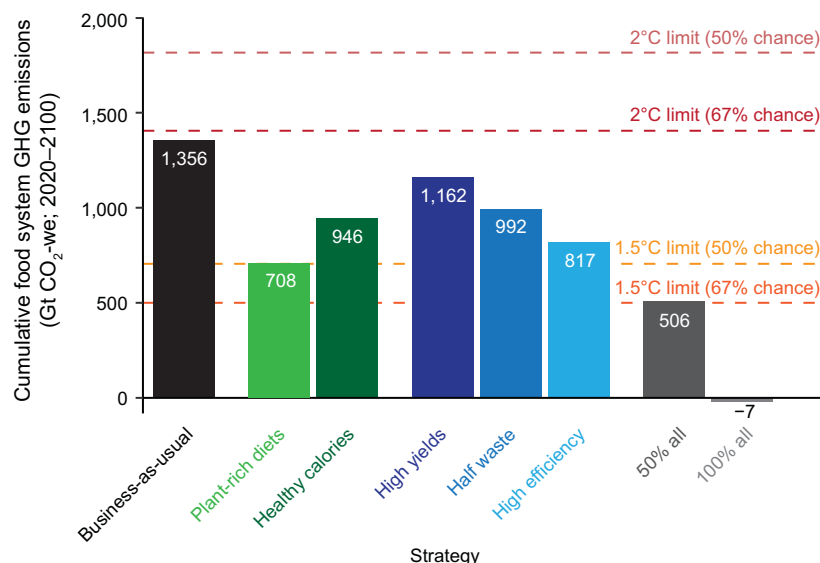


Fig. 1. Projected cumulative 2020 to 2100 GHG emissions solely from the global food system for business-as-usual emissions and for various food system changes that lead to emission reductions.

Food system changes are gradually adopted between 2020 and 2050. Bars are colored by type of change to the food system: The black bar indicates business-as-usual emissions, green bars indicate changes to dietary patterns, blue bars indicate changes to food supply chains, and gray bars indicate combined changes of all five individual strategies. The plant-rich diet scenario is based on EAT-Lancet recommendations (15), the healthy calorie scenario contains ~2100 daily kilocalories per person, the high yields scenario involves yields that are 50% above current maximum potential yields, the half waste scenario has food loss and waste reduced by 50%, and the high efficiency scenario indicates a 40% reduction in GHG emissions per unit of food produced. The two rightmost columns indicate a global transition halfway (50% all) or entirely (100% all) to adoption by 2050 of all five strategies: plant-rich diet, healthy calories, high yields, half waste, and high efficiency changes. Horizontal lines indicate the maximum cumulative emissions from all sources (food and nonfood) compatible with a 50 or 67% likelihood of achieving the 2° (red) and 1.5°C (orange) temperature targets.

and meat, such as a Mediterranean diet or planetary health diet (15)]; (ii) adjusting global per capita caloric consumption to healthy levels; (iii) achieving high yields by closing yield gaps and improving crop genetics and agronomic practices; (iv) reducing food loss and waste by 50%; and (v) reducing the GHG intensity of foods by increasing the efficiency of production, such as by altering management regimes (e.g., precise use of nitrogen fertilizer and other inputs) or technological implementation (e.g., additives to ruminant feed). We also explore the potential GHG benefits of partial (50%) or complete (100%) adoption of all five strategies simultaneously. Other combinations of strategies and their levels of adoption are provided in data S2. Although we discuss food system transitions at the global scale, the magnitude and direction of the transitions will vary by country.

We find that cumulative food system GHG emissions from 2020 to 2100 can be reduced by 14 to 48% through changes in dietary composition and healthier caloric consumption, through increased crop yields, through decreased food loss and waste, or through increased emissions efficiency of food production,

provided that these strategies are adopted individually and gradually such that they are fully adopted by 2050 (Fig. 1). If all five strategies were to be partially implemented together (50% adoption of each), cumulative emissions through 2100 could be reduced by 63% relative to business-as-usual. Full adoption of all five strategies could result in a food system with marginally negative net cumulative emissions because of lowered emissions and net carbon sequestration on abandoned croplands (Fig. 1).

GHG emissions from all human activities affect global climate. As such, to meet a given emissions limit, there is a tradeoff between food and nonfood emissions within a total cumulative budget: Higher emissions from the global food system necessitate lower emissions from other sectors, and vice versa. To illustrate how emissions from all human activities might be kept under the emissions limits, we consider them in the context of an increasingly decarbonized future in which all nonfood emissions and all food-related emissions from fossil fuel combustion decline linearly from current levels to zero by 2050 (4). This rate of reduction is approximately in line with the rates of decar-

bonization estimated to be needed to meet the 1.5°C target in global integrated assessment models (16). We find that in this increasingly decarbonized future, total global emissions from all sources (business-as-usual food plus nonfood) would exceed the 1.5°C limit within 11 years, and they would exceed the 2°C limit before the end of the century (Fig. 2A).

Assuming this linear reduction to decarbonization in 2050, meeting the 2°C target is plausible through the use of numerous food system strategies, provided that they are also adopted by 2050 (Fig. 2A). As is well known, dietary changes—such as the adoption of plant-rich diets—can greatly reduce emissions (5, 6, 17). Even in the absence of dietary changes, achieving either high yields, high agricultural efficiency, or a 50% reduction in food waste alone could also meet the 2°C limit, as could partial achievement of various strategies (Fig. 2A).

Meeting the 1.5°C target with this linear decarbonization by 2050 requires at least partial achievement of multiple food system strategies: None of the five individual strategies alone are sufficient. If full implementation of these food and nonfood emission changes were to be delayed by 25 years to 2075, then even 100% adoption of all five strategies would preclude meeting the 1.5°C target (Fig. 2B). For this case of slower implementation, the 2°C target could be met only by at least a 50% adoption of all five strategies, and not by any single strategy (Fig. 2B). This is because a slower adoption of food system strategies, a slower reduction of fossil fuel use in the food system, and a slower reduction in nonfood emissions each necessitates larger changes to the food system to meet targets.

The need for rapid reduction in GHG emissions from fossil fuels to meet the 1.5° or 2°C targets is widely acknowledged. We show that the same is true for food systems: Even if fossil fuel emissions were rapidly reduced, emissions from the global food system are on a trajectory that would prevent achievement of the 1.5° and 2°C targets before the end of the century. Our analyses also suggest there are many opportunities to meet the 1.5° or 2°C emission targets. Previous analyses have suggested that global food system emissions might increase by up to 80% from 2010 to 2050 (5, 6, 17). Our findings—consistent with these results (see supplementary materials)—improve on these forecasts by explicitly linking food systems to IPCC cumulative emissions limits (13), using a reporting method to include CH₄ in this framework (14), increasing the breadth of scenarios analyzed, allowing for different levels and different rates at which food system transitions occur, providing annual emissions estimates, and forecasting beyond 2050 to 2100.

We show that meeting the 1.5° and 2°C targets will likely require extensive and unprecedented changes to the global food system.

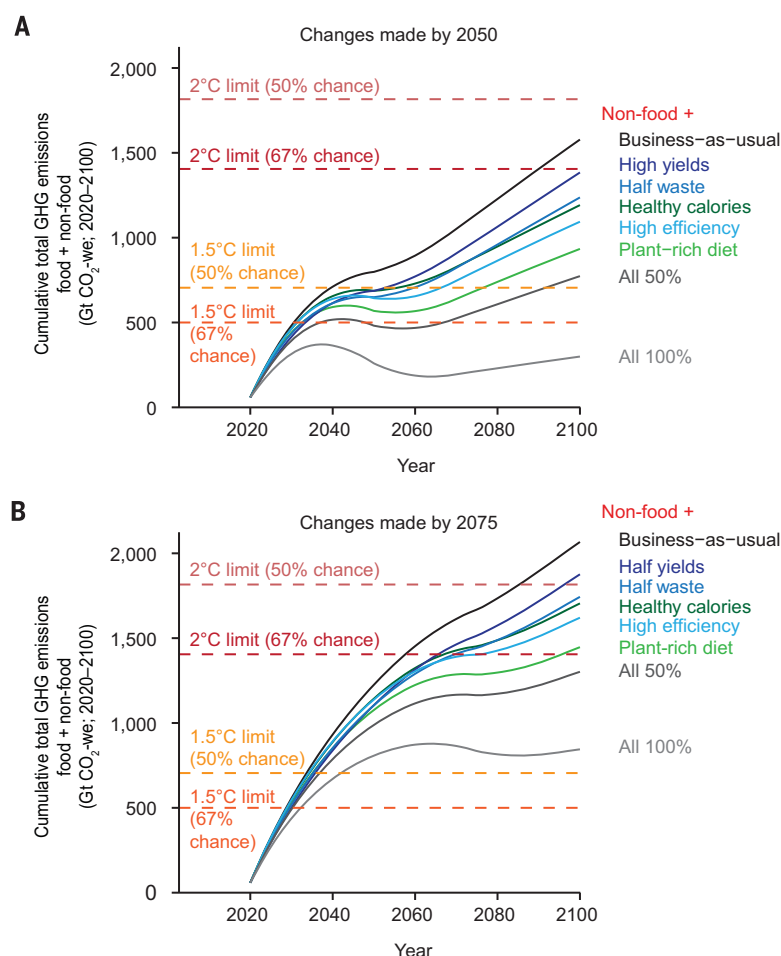


Fig. 2. Estimated GHG emissions from all human activities (food plus nonfood) for different food system changes and different rates of emissions reductions from fossil fuels and food systems. (A and B) Nonfood emissions are linearly reduced to zero from 2020 to 2050 (A) or from 2020 to 2075 (B). Solid curves show cumulative emissions from all human activities if different food system strategies were to be implemented. Fossil fuel emissions from within the food system are also assumed to be reduced at the same rate as for emissions from outside the food system. Horizontal dashed lines indicate maximum cumulative emissions from all sources (food and nonfood) compatible with a 50 or 67% likelihood of meeting the remaining 2° (red) and 1.5° (orange) temperature targets.

Recent studies have provided insight into plausibly achievable ways to reduce food system GHG emissions. Large-scale field trials in China and the United States have shown that changes in farm management could reduce nitrogen fertilizer use and its associated GHG emissions while increasing farmer profits (18, 19). Rapid increases in crop yields that decrease land clearing and its emissions have been achieved through access to improved seeds and fertilizers (20). Such increases in yields might also be achieved through the adoption of agroecological production practices—including cover crops, integrated pest management, and increased use of precision agriculture (21, 22)—but will require different management interventions in different regions (23). Food awareness, reformulation, and labeling; changes in the food environment; and education and aware-

ness campaigns have shifted consumer food purchases in numerous countries (24, 25). Carbon taxation might also be effective (26). Food loss and waste could be reduced by improvements to infrastructure, such as grain storage and refrigeration, or by innovative methods to sell food that would otherwise be wasted (9). Food system changes that reduce GHG emissions may offer additional benefits (27), including progress toward targets set in the United Nations' Sustainable Development Goals (28), such as decreased nutrient pollution (6), reduced water pollution and scarcity (6), decreased land-use change (5, 6, 17), improved biodiversity outcomes (29), and, if dietary composition and caloric consumption are improved, reduced prevalence of obesity, diabetes, heart disease, and premature mortality (30).

Time is of the essence in addressing GHG emissions. Any delays will necessitate more ambitious and expeditious implementation of emissions reduction strategies if global temperature targets are to be met. We show that there are many opportunities to keep emissions from food systems and other activities within the global emissions limits for the 1.5° and 2° targets. The global challenge of finding and implementing feasible, ethical, and equitable policies to reduce net GHG emissions will require the rapid adoption of coordinated solutions, both within and outside of the food system, that are tailored to the needs and customs of different countries and the communities within them.

REFERENCES AND NOTES

1. J. Rogelj, D. Shindell, K. Jiang, S. Ffifita, P. Forster, V. Ginzburg, C. Handa, H. Kheshgi, S. Kobayashi, E. Kriegler, L. Mundaca, R. Séférian, M. V. Vilarinho, in *Global Warming of 1.5°C: An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty*, V. Masson-Delmotte, P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P. R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J. B. R. Matthews, Y. Chen, X. Zhou, M. I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, T. Waterfield, Eds. [Intergovernmental Panel on Climate Change (IPCC), 2018], pp. 93–174.
2. S. Vermeulen, B. M. Campbell, J. S. I. Ingram, *Annu. Rev. Environ. Resour.* **37**, 195–222 (2012).
3. C. Rosenzweig et al., *Nat. Food* **1**, 94–97 (2020).
4. J. Gütschow, L. Jeffery, R. Gieseke, The PRIMAP-hist national historical emissions time series (1850–2016), version 2.0., GFZ Data Services (2019); <https://doi.org/10.5880/pik.2019.001>.
5. D. Tilman, M. Clark, *Nature* **515**, 518–522 (2014).
6. M. Springmann et al., *Nature* **562**, 519–525 (2018).
7. United Nations Department of Economic and Social Affairs (UN DESA), "World Population Prospects: The 2017 Revision" (2017); www.un.org/development/desa/publications/world-population-prospects-the-2017-revision.html.
8. M. Clark, D. Tilman, *Environ. Res. Lett.* **12**, 064016 (2017).
9. J. Gustavsson, C. Cederberg, U. Sonesson, R. van Otterdijk, A. Meybeck, *Global Food Losses and Food Waste: Extent, Causes and Prevention* (Food and Agricultural Organization, 2011).
10. J. Poore, T. Nemecek, *Science* **360**, 987–992 (2018).
11. A. Ruesch, H. Gibbs, "New IPCC Tier-1 Global Biomass Carbon Map for the Year 2000" (Carbon Dioxide Information Analysis Center, 2008); <http://cdiac.ess-dive.lbl.gov>.
12. T. Hengl et al., *PLOS ONE* **12**, e0169748 (2017).
13. P. Forster, D. Huppmann, E. Kriegler, L. Mundaca, C. Smith, J. Rogelj, R. Séférian, in *Global Warming of 1.5°C: An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty*, V. Masson-Delmotte, P. Zhai, H.-O. Pörtner, D. Roberts, J. Skea, P. R. Shukla, A. Pirani, W. Moufouma-Okia, C. Péan, R. Pidcock, S. Connors, J. B. R. Matthews, Y. Chen, X. Zhou, M. I. Gomis, E. Lonnoy, T. Maycock, M. Tignor, T. Waterfield, Eds. (IPCC, 2018), pp. 2SM-1–2SM-50.
14. M. Cain et al., *npj Clim. Atmos. Sci.* **2**, 29 (2019).
15. W. Willett et al., *Lancet* **393**, 447–492 (2019).
16. S. Roe et al., *Nat. Clim. Chang.* **9**, 817–828 (2019).
17. B. Bajželj et al., *Nat. Clim. Chang.* **4**, 924–929 (2014).
18. Z. Cui et al., *Nature* **555**, 363–366 (2018).
19. S. Sela et al., *Agron. J.* **108**, 1726–1734 (2016).
20. Z. Druilhe, J. Barreiro-Huélé, "Fertilizer subsidies in sub-Saharan Africa" [ESA Working Paper 288997, Food and Agriculture Organization of the United Nations, Agricultural Development Economics Division (ESA), 2012]; <https://doi.org/10.22004/ag.econ.288997>.

21. H. C. J. Godfray, T. Garnett, *Phil. Trans. R. Soc. B* **369**, 20120273 (2014).
22. G. Philip Robertson *et al.*, *Bioscience* **64**, 404–415 (2014).
23. P. Pradhan, G. Fischer, H. van Velthuis, D. E. Reusser, J. P. Kropp, *PLOS ONE* **10**, e0129487 (2015).
24. M. A. Colchero, J. Rivera-Dommarco, B. M. Popkin, S. W. Ng, *Health Aff.* **36**, 564–571 (2017).
25. J. Ranganathan, D. Vennard, R. Waite, B. Lipinski, T. Searchinger, P. Dumas, A. Forslund, H. Guyomard, S. Manceron, E. Marajo Petitzon, C. Mouél, P. Havlik, M. Herrero, X. Zhang, S. Wirseniuss, F. Ramos, X. Yan, M. Phillips, R. Mungkung, “Shifting Diets for a Sustainable Future” (Installment 11 of Creating a Sustainable Food Future, World Resources Institute, 2016); www.wri.org/publication/shifting-diets.
26. M. Springmann *et al.*, *Nat. Clim. Chang.* **7**, 69–74 (2016).
27. C. Mbow, C. Rosenzweig, L. G. Barioni, T. G. Benton, M. Herrero, M. Krishnapillai, E. Liwenga, P. Pradhan, M. G. Rivera-Ferre, T. Sapkota, F. N. Tubiello, Y. Xu, in *Climate Change and Land: An IPCC Special Report on climate change, desertification, land degradation, sustainable land management, food security, and greenhouse gas fluxes in terrestrial ecosystems*, P. R. Shukla, J. Skea, E. Calvo Buendia, V. Masson-Delmotte, H.-O. Pörtner, D. C. Roberts, P. Zhai, R. Slade, S. Connors, R. van Diemen, M. Ferrat, E. Haughey, S. Luz, S. Neogi, M. Pathak, J. Petzold, J. Portugal Pereira, P. Vyas, E. Huntley, K. Kissick, M. Belkacemi, J. Malley, Eds. (IPCC, 2019), pp. 437–550.
28. United Nations General Assembly, “Transforming Our World: The 2030 Agenda for Sustainable Development” (UN, 2015); <https://sustainabledevelopment.un.org/post2015/transformingourworld/publication>.
29. D. Tilman *et al.*, *Nature* **546**, 73–81 (2017).
30. M. Springmann *et al.*, *Lancet Planet. Health* **2**, e451–e461 (2018).

ACKNOWLEDGMENTS

We are grateful to M. Springmann, C. Godfray, P. Scarborough, and K. Kimmel, who provided comments on the manuscript.

Funding: This research was made possible through funding from the Balzan Award Prize to D.T.; the Grand Challenges Research Initiative at the University of Minnesota; the Wellcome Trust, Our Planet Our Health (Livestock, Environment and People), award no. 205212/Z/16/Z; and the U.S. Department of Agriculture (MIN-12-083). This publication was also developed as part of the Center for Air, Climate, and Energy Solutions, which was supported under Assistance Agreement no. R835873 awarded by the U.S. Environmental Protection Agency (EPA). It has not been formally reviewed by the EPA. The views expressed in this document are

solely those of the authors and do not necessarily reflect those of the EPA. The EPA does not endorse any products or commercial services mentioned in this publication. **Author contributions:** J.D.H., N.G.G.D., K.C., S.K.T., D.T., M.A.C., and J.L. conceptualized the project; M.A.C., J.D.H., D.T., N.G.G.D., and J.L. developed the project methodology; M.A.C., N.G.G.D., D.T., and J.D.H. performed the analysis; M.A.C., J.D.H., and D.T. visualized the results; M.A.C. wrote the original draft; and all authors contributed to the review and editing of the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Data used to create the figures is available in data S1. Results from all analyses are available in data S2. All data inputs and code required to run the analyses are available for download as data S3.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6517/705/suppl/DC1
Materials and Methods
Supplementary Text
References (31–60)
Data S1 to S3

8 January 2020; accepted 1 September 2020
10.1126/science.aba7357

BATTERIES

Rational design of layered oxide materials for sodium-ion batteries

Chenglong Zhao^{1,2*}, Qidi Wang^{3,4*}, Zhenpeng Yao^{5*}, Jianlin Wang⁶, Benjamín Sánchez-Lengeling⁵, Feixiang Ding^{1,2}, Xingguo Qi^{1,2}, Yaxiang Lu^{1,2†}, Xuedong Bai⁶, Baohua Li³, Hong Li^{1,2}, Alán Aspuru-Guzik^{5,7†}, Xuejie Huang^{1,2}, Claude Delmas^{8†}, Marnix Wagemaker^{9†}, Liquan Chen¹, Yong-Sheng Hu^{1,2,10†}

Sodium-ion batteries have captured widespread attention for grid-scale energy storage owing to the natural abundance of sodium. The performance of such batteries is limited by available electrode materials, especially for sodium-ion layered oxides, motivating the exploration of high compositional diversity. How the composition determines the structural chemistry is decisive for the electrochemical performance but very challenging to predict, especially for complex compositions. We introduce the “cationic potential” that captures the key interactions of layered materials and makes it possible to predict the stacking structures. This is demonstrated through the rational design and preparation of layered electrode materials with improved performance. As the stacking structure determines the functional properties, this methodology offers a solution toward the design of alkali metal layered oxides.

Integration of intermittent renewable energy sources demands the development of sustainable electrical energy storage systems (1). Compared with lithium (Li)-ion batteries, the abundance and low cost of sodium (Na) make Na-ion batteries promising for smart grids and large-scale energy storage applications (2, 3). Li-ion layered oxides, with

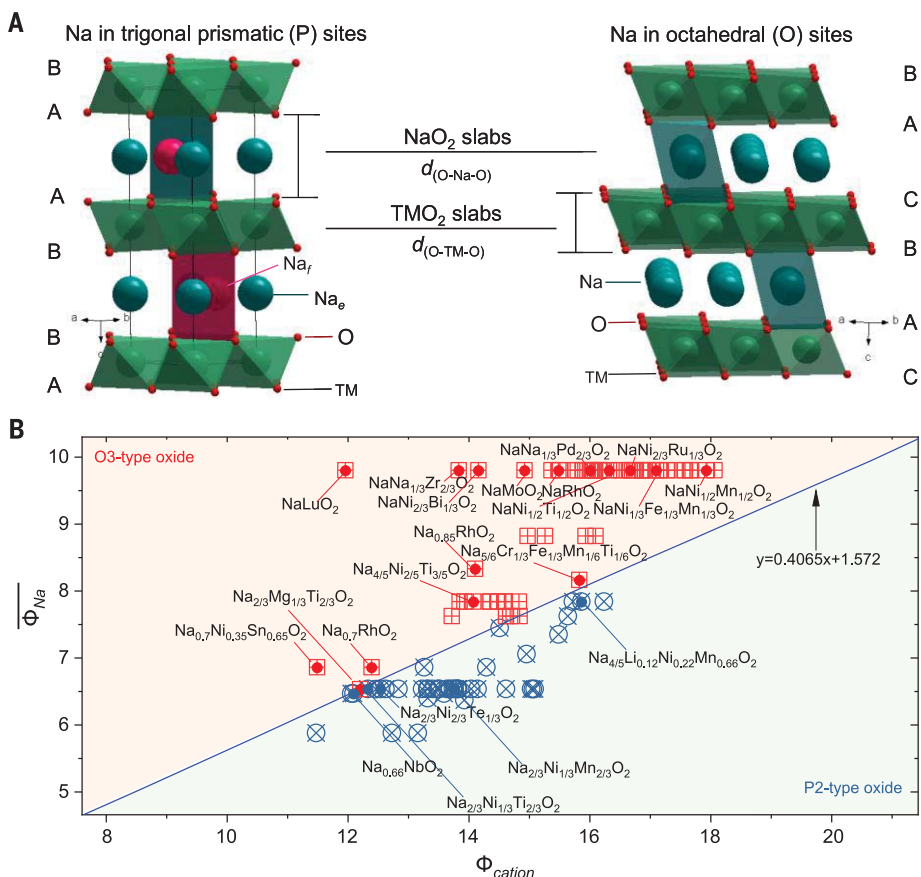
the general formula LiTMO_2 , have represented the dominant family of electrode materials for Li-ion batteries since 1980 (4). TM stands for one or multiple transition metal elements that facilitate the redox reaction associated with Li-ion (de-)intercalation. The layered structures are built up by edge-sharing TMO_6 octahedra, forming repeating layers between which

Li ions are positioned in the octahedral (O) oxygen environment, leading to the so-called O-type stacking. The structure offers high compositional diversity, providing tunable electrochemical performance, of which well-known examples are LiCoO_2 and Ni-rich $\text{LiNi}_{1-y}\text{Co}_y\text{Mn}_{1-y-z}\text{O}_2$. In search of electrodes for Na-ion batteries, layered oxides (Na_xTMO_2) offered the natural starting point (5). However, a key difference is that for Na-ion oxides, in addition to O-type, P-type stacking can occur, in which P-type refers to prismatic Na-ion coordination (Fig. 1A). These stackings show distinctly different electrode performance, in which the most studied layered stacking configurations are P2 and O3 types (Fig. 1A), referring to the ABBA and ABCABC oxygen stacking, respectively (6). P2-type oxides usually provide higher Na-ion conductivity and better structural integrity against the O3 analogs, which is responsible for the high power density and good cycling stability (7). However, the lower initial Na content of P2-type electrodes limits the storage capacity in the first charge compared with high-Na-content O3-type materials (8). Usually, the structural transition between the O- and P-type can occur upon Na-ion (de)intercalation during (dis)charging, typically degrading cycle stability (2, 3).

In a search for electrodes with good chemical and dynamic stability and high Na storage

Fig. 1. Cationic potential and its use in Na-ion layered oxides.

(A) Schematic illustration of crystal representative P2-type (hexagonal) and O3-type (rhombohedral) layered oxides. (B) Cationic potential of representative P2- and O3-type Na-ion layered oxides, considering the Na content, oxidation state of transition metals, and TM composition (see supplementary text and fig. S3 for details).



performance, various P2- and O3-type Na-ion layered oxides have been synthesized and investigated (9, 10). However, effective guidelines toward the design and preparation of optimal electrode materials are lacking. Crystal structures of P2- and O3-type layered oxides can be differentiated on the basis of the ratio between the interlayer distance of the Na metal layer $d_{(\text{O-Na-O})}$ and the TM layer distance $d_{(\text{O-TM-O})}$ (11), in which a ratio of ~ 1.62 distinguishes P2- and O3-type oxides (fig. S1 and table S1) (12). The larger ratio of P2-type oxides originates from the more localized electron distribution within the TMO_2 slabs, which results in a weaker repulsion between the adjacent NaO_2 slabs and consequentially a stronger repulsion between the adjacent TMO_2 slabs. This hints that the electron distribution plays an important role in the competition between the P- and O-type stackings in layered oxides.

Ionic potential (Φ) is an indicator of the charge density at the surface of an ion, which is the ratio of the charge number (n) with the ion radius (R) introduced by G. H. Cartledge (13), reflecting the cation polarization power. The ionic potential shows the expected increase with oxidation state and atom mass (fig. S2 and table S2), a consequence of the less localized orbitals.

Aiming at a simple descriptor for layered oxides, we express the extent of the cation electron density and its polarizability, normalized to the ionic potential anion (O), by defining the “cationic potential”:

$$\Phi_{\text{cation}} = \frac{\Phi_{\text{TM}} \Phi_{\text{Na}}}{\Phi_{\text{O}}} \quad (1)$$

where Φ_{TM} represents the weighted average ionic potential of TMs, defined as $\Phi_{\text{TM}} = \sum \frac{w_i n_i}{R_i}$; w_i is the content of TM_i having charge number n_i and radius R_i ; and Φ_{Na} represents the weighted average ionic potential of Na defined as $\Phi_{\text{Na}} = \frac{x}{R_{\text{Na}}}$. Charge balance in Na_xTMO_2

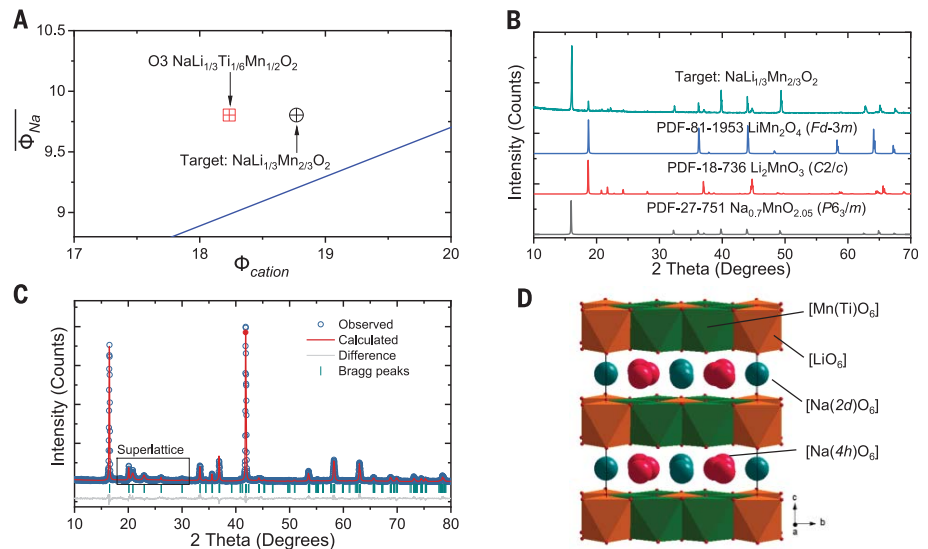


Fig. 2. Designing an O3-type oxide. (A) Analysis of the cationic potential of Na-Li-Mn(Ti)-O oxides (see tables S8 and S9 for details). (B) XRD patterns of the targeted $\text{NaLi}_{1/3}\text{Mn}_{2/3}\text{O}_2$ and the standard references. (C) Rietveld refinement of XRD pattern of $\text{NaLi}_{1/3}\text{Ti}_{1/6}\text{Mn}_{1/2}\text{O}_2$ (see tables S10 to S12 for details). (D) Schematic illustration of the corresponding structure with the Li/Mn(Ti) ordering in the $[\text{Li}_{1/3}\text{Ti}_{1/6}\text{Mn}_{1/2}]_2\text{O}_2$ slabs.

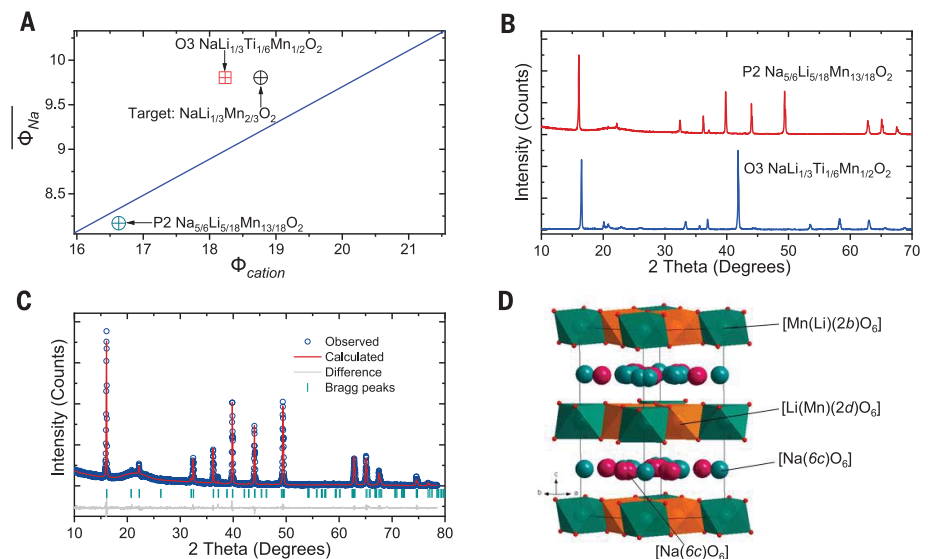


Fig. 3. Designing a P2-type oxide. (A) Analysis of cationic potential of Na-Li-Mn-O oxides (see tables S13 and S14 for details). (B) XRD patterns of $\text{NaLi}_{1/3}\text{Ti}_{1/6}\text{Mn}_{1/2}\text{O}_2$ and $\text{Na}_{5/6}\text{Li}_{5/18}\text{Mn}_{13/18}\text{O}_2$ oxides. (C) Rietveld refinement of XRD pattern of $\text{Na}_{5/6}\text{Li}_{5/18}\text{Mn}_{13/18}\text{O}_2$ (see tables S15 to S17 for details). (D) Schematic illustration of the corresponding structure with the Li/Mn ordering in the $[\text{Li}_{5/18}\text{Mn}_{13/18}]\text{O}_2$ slabs.

composition demands $\sum w_i n_i = 4 - x$, where x represents Na content and 4 is the total oxidation state to charge compensate O^{2-} .

The cationic potential Φ_{cation} versus the weighted average Na ionic potential Φ_{Na} of reported P2- and O3-type layered oxides results in the phase map shown in Fig. 1B. The distinct P2- and O3-type regions indicate that the cationic potential is an accurate descriptor of the interslab interaction and, thereby, the structural competition between P2- and O3-type structures.

A larger cationic potential (Eq. 1) implies stronger TM electron cloud extend and interlayer electrostatic repulsion resulting in the P2-type structure, with more covalent TM-O bonds and an increased $d_{(\text{O-Na-O})}$ distance (fig. S4). Opposing this, a larger mean Na ionic potential, achieved by increasing Na content, increases the shielding of the electrostatic repulsion between the TMO_2 slabs, favoring the O3-type structure.

The phase map (Fig. 1B) shows that very small differences in TM or Na content can

¹Key Laboratory for Renewable Energy, Beijing Key Laboratory for New Energy Materials and Devices, Beijing National Laboratory for Condensed Matter Physics, Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China. ²Center of Materials Science and Optoelectronics Engineering, University of Chinese Academy of Sciences, Beijing 100049, China. ³Shenzhen Key Laboratory on Power Battery Safety and Shenzhen Geim Graphene Center, School of Shenzhen International Graduate, Tsinghua University, Guangdong 518055, China. ⁴School of Materials Science and Engineering, Tsinghua University, Beijing 100084, China. ⁵Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA. ⁶State Key Laboratory for Surface Physics, Institute of Physics, Chinese Academy of Sciences, Beijing, 100190, China. ⁷Department of Chemistry and Department of Computer Science, University of Toronto, Toronto, Ontario M5S 3H6, Canada. ⁸Université de Bordeaux, Bordeaux INP, ICMCB UMR 5026, CNRS, 33600 Pessac, France. ⁹Department of Radiation Science and Technology, Delft University of Technology, Mekelweg 15, 2629JB Delft, Netherlands. ¹⁰Yangtze River Delta Physics Research Center, Liyang 213300, China. *These authors contributed equally to this work. †Corresponding author. Email: yxlu@iphy.ac.cn (Y.L.); aspu@utoronto.ca (A.A.-G.); delmas@cmcb-bordeaux.cnrs.fr (C.D.); m.wagemaker@tudelft.nl (M.W.); yshu@iphy.ac.cn (Y.S.-H.)

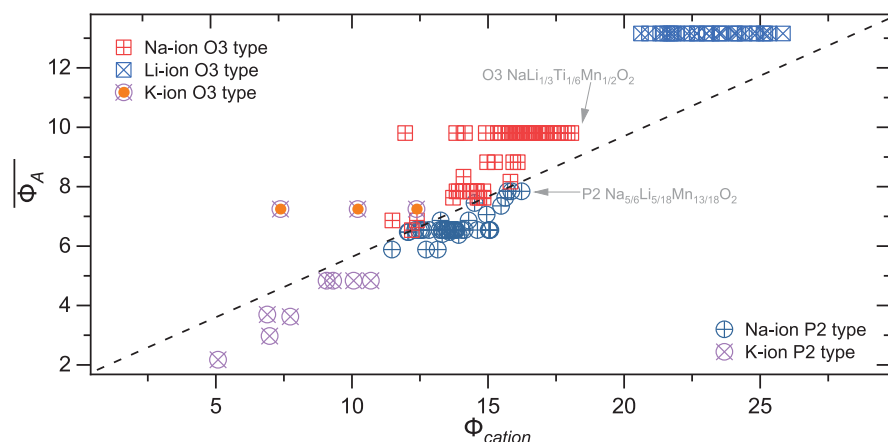


Fig. 4. Cationic potential phase map for alkali metal layered oxides. Summary of reported alkali metal layered materials including Li-/Na-/K-ion oxides (see tables S18 and S19 for details).

result in a transition between P2- and O3-type structures. To illustrate this, we consider layered oxides with the composition $\text{Na}_{2/3}\text{TMO}_2$, which typically crystallizes in a P2-type structure for the low Na content, such as P2- $\text{Na}_{2/3}\text{CoO}_2$ (14) or P2- $\text{Na}_{2/3}\text{Ni}_{1/3}\text{Ti}_{2/3}\text{O}_2$ (15). However, replacing Ni^{2+} with Mg^{2+} in P2- $\text{Na}_{2/3}\text{Ni}_{1/3}\text{Ti}_{2/3}\text{O}_2$, facilitated by their similar ionic radii (16), leads to $\text{Na}_{2/3}\text{Mg}_{1/3}\text{Ti}_{2/3}\text{O}_2$, for which the cationic potential predicts the O3-type structure, which is difficult to predict even with complex electrostatic energy calculations (15). In this case, the smaller ionic potential of Mg^{2+} against Ni^{2+} (Fig. 1B) decreases Φ_{cation} ; the resulting lower covalence of Mg/Ti-O bonds increases the charge carried by the oxygens and thereby weakens the repulsion between the TM layers, resulting in an O3-type structure (fig. S5, A and B, and tables S6 and S7). Substituting $1/6 \text{ Mg}^{2+}$ by Ni^{2+} in $\text{Na}_{2/3}\text{Mg}_{1/3}\text{Ti}_{2/3}\text{O}_2$ to $\text{Na}_{2/3}\text{Ni}_{1/6}\text{Mg}_{5/6}\text{Ti}_{2/3}\text{O}_2$ moves it back into a P2-type structure (fig. S5B), illustrating how near these compositions are to the line separating the P2- and O3-type phases. Several other examples demonstrating that the proposed cationic potential approach captures the subtle balance between the P2- and O3-type layered Na_xTMO_2 structures are provided in the supplementary text and figs. S5C and S6.

Delmas *et al.* (6, 17) used the Rouxel diagram (18) to distinguish Na_xTMO_2 stacking structures, demonstrating that both Na content and the ionicity and covalence of bonds are the important factors. However, this method only accounts for the difference in Pauling's electronegativity (fig. S7 and table S4), which makes it impossible to predict the structure of oxides with the same TMs in different oxidation states (6, 17) (e.g., Mn^{4+} and Mn^{3+} in $\text{Na}_{0.7}\text{MnO}_2$) or for multiple-component systems (see supplementary text, fig. S8, and table S5 for details). The cationic potential correctly predicts the stacking structure for these cases, providing a guideline for the development of Na-ion layered oxides.

Using the cationic potential as a guide, we design specific stacking structures by controlling the Na content and TM composition. A notable starting point is $\text{NaLi}_{1/3}\text{Mn}_{2/3}\text{O}_2$, the analog of $\text{LiLi}_{1/3}\text{Mn}_{2/3}\text{O}_2$ (Li_2MnO_3), providing capacity on the basis of oxygen redox chemistry. This composition has not been prepared so far, although theoretical calculations argue $\text{NaLi}_{1/3}\text{Mn}_{2/3}\text{O}_2$ is stable in an O3-type structure (19). Various experimental conditions were attempted to prepare this composition in an O3-type structure, but always a P2-type component, in addition to other phases, was obtained. Lowering the cationic potential suggests that a possible route to prepare the O3-type structure is partial substitution of Mn^{4+} by Ti^{4+} (Fig. 2A), where Ti^{4+} has a lower ionic potential. $\text{NaLi}_{1/3}\text{Ti}_{1/6}\text{Mn}_{1/2}\text{O}_2$ was successfully prepared in the predicted O3-type structure by a typical solid-state reaction (materials and methods). Notably, $\text{NaLi}_{1/3}\text{Mn}_{2/3}\text{O}_2$ could not be synthesized as an O3-type structure by using the same method (Fig. 2B). Rietveld refinement of the x-ray diffraction (XRD) pattern confirmed the layered rock-salt structure (Fig. 2C), in which the NaO_2 layers alternate with the mixed $[\text{Li}_{1/3}\text{Ti}_{1/6}\text{Mn}_{1/2}]\text{O}_2$ slabs (Fig. 2D). The $(1/3, 1/3, 1)$ superstructure peaks in 20° to 30° suggest Li/Mn(Ti) ordering in a honeycomb pattern, which is also confirmed by the aberration-corrected scanning transmission electron microscopy (fig. S9). This ordered arrangement of Li and Mn(Ti) in the TMO_2 slabs has not been observed in O3-type Na-ion oxides with exclusively 3d TMs. The electrochemical properties (see supplementary text and fig. S10A for details) demonstrate an energy density of ~ 630 watt-hours (Wh) kg^{-1} , higher than that of the reported O3-type electrodes.

We then use cationic potential to design a P2-type structure aiming at an anomalous high Na-content of $x > 0.67$, again starting from $\text{NaLi}_{1/3}\text{Mn}_{2/3}\text{O}_2$. To avoid formation of

an O3-type structure, the dividing line in Fig. 1B demonstrates that we should increase the cationic potential (Eq. 1), assuming that Na content remains constant, which can be realized by increasing the ionic potential at TM sites. On the basis of the cationic potential, a P2-type structure with $x = 1$ ($\Phi_{\text{Na}} = 9.8$) will demand an extremely large TM ionic potential (larger than that of Mn^{4+} , having the largest value among the widely used TMs). Therefore, the Na content in $\text{NaLi}_{1/3}\text{Mn}_{2/3}\text{O}_2$ should be lowered, which can be achieved through charge compensation by decreasing the Li and increasing the Mn content. Following this route, the cationic potential predicts that $\text{Na}_{5/6}\text{Li}_{5/18}\text{Mn}_{13/18}\text{O}_2$ composition with high Na content should have the P2-type structure (Fig. 3A), which was indeed successfully prepared (Fig. 3B). So far, layered oxides prepared with such high Na content usually crystallize as an O3-type structure. Compared with the O3-type $\text{NaLi}_{1/3}\text{Ti}_{1/6}\text{Mn}_{1/2}\text{O}_2$, the (002) peak of the P2-type structure shifts toward lower diffraction angles, indicating the expected increase in the c axis of the unit cell (Fig. 3B). Rietveld refinement of the XRD pattern reveals that this P2-type layered structure can be indexed in the hexagonal $P6_3$ space group (Fig. 3, C and D). The electron energy loss spectroscopy mapping reveals a uniform distribution of the Na, Mn, and O elements in the platelike particles (fig. S11). This as-prepared high-Na-content material has a considerably higher capacity of >200 milliamperes-hours (mAh) g^{-1} (fig. S10B).

Extending the cationic potential to other alkali metal layered oxides, Li ion (fig. S12) and K ion (fig. S13), results in phase maps shown in Fig. 4. The cationic potential (Eq. 1) is found to increase from K to Na to Li ion owing to the increasing ability to shield the TMO_2 interslab interaction. As a consequence, K_xTMO_2 mainly crystallizes as the P2-type and Li_xTMO_2 as the O3-type structure, whereas Na_xTMO_2 is the most notable family, as the shielding strength is at the tipping point between P2- and O3-type structures. The distribution of reported layered electrodes exhibits a clear trend by clustering around the dividing line (Fig. 4). For more than 100,000 new compositions, up to quaternary compositions on the TM position, the cationic potential is used to calculate the most stable stacking structure, resulting in a distribution of compositions in the phase map around the dividing line (see figs. S14 and S15 and supplementary text for details). This demonstrates how the cationic potential can be used to predict the structure of new Na_xTMO_2 layered materials, on the basis of specific compositional demands. It is worth noting that the other parts far away from the line may also lead to other types of TM-oxide phases (e.g., rock salt, spinel), or may not lead to stable structures at all, which is the subject of ongoing investigations.

In summary, the ionic potential is a measure of the polarization of ions, mainly reflecting the influence of electrostatic energy on the system. Because the main difference between P- and O-type structures is the electrostatic polarization between AO_2 (A, alkali metals) and TMO_2 slabs, we can apply the proposed cationic potential method to distinguish and design materials, especially useful for Na-ion layered oxides. For entropy-dominated phases, disordered compounds resulting from mechanical milling (20), or oxides prepared under particular conditions (21, 22), metastable structures, or nonequilibrium phases (23), as well as the local distortion of TMs (e.g., due to Jahn-Teller effect on Mn^{3+}), the cationic potential approach does not provide a sensible guideline. Moreover, the cationic potential only predicts whether the proposed material will crystallize in a P- or O-type structure, and one composition has only one structure. The actual obtained phases depend strongly on the nature of precursors and the conditions/atmosphere of thermal treatment, among others, which may cause the difference in stoichiometry and dynamic process, leading to structural changes. Further structural information is required to decide whether the corresponding material is stable and/or synthesizable in practice and calls for extensive investigation. Additionally, prediction of stacking structures is challenging for density functional theory methods because of the difficulty of predicting the localized nature of TM orbitals, especially for complicated TM compositions that have an enormous configurational space. We demonstrated the use of cationic potentials to tune the TMO_2 interslab interaction, contributing

to the important categories of layered materials. The currently known layered materials are either low-Na-content ($x = 2/3$) P2-type oxides or high-Na-content ($x = 1$) O3-type oxides; we suggest further exploration of high-Na-content P2-type oxides and low-Na-content O3-type oxides through the as-proposed cationic potential.

REFERENCES AND NOTES

1. B. Dunn, H. Kamath, J.-M. Tarascon, *Science* **334**, 928–935 (2011).
2. S.-W. Kim, D.-H. Seo, X. Ma, G. Ceder, K. Kang, *Adv. Energy Mater.* **2**, 710–721 (2012).
3. M. H. Han, E. Gonzalo, G. Singh, T. Rojo, *Energy Environ. Sci.* **8**, 81–102 (2015).
4. K. Mizushima, P. C. Jones, P. J. Wiseman, J. B. Goodenough, *Mater. Res. Bull.* **15**, 783–789 (1980).
5. N. Yabuuchi *et al.*, *Nat. Mater.* **11**, 512–517 (2012).
6. C. Delmas, C. Fouassier, P. Hagenmuller, *Physica B+C* **99**, 81–85 (1980).
7. C. Fouassier, C. Delmas, P. Hagenmuller, *Mater. Res. Bull.* **10**, 443–449 (1975).
8. S. Komaba *et al.*, *Inorg. Chem.* **51**, 6211–6220 (2012).
9. J. M. Paulsen, R. A. Donabarger, J. R. Dahn, *Chem. Mater.* **12**, 2257–2267 (2000).
10. J. Billaud *et al.*, *Energy Environ. Sci.* **7**, 1387–1391 (2014).
11. M. Guilmard, L. Croguennec, C. Delmas, *J. Power Sources* **150**, A1287–A1293 (2003).
12. C. Zhao, M. Avdeev, L. Chen, Y.-S. Hu, *Angew. Chem. Int. Ed.* **57**, 7056–7060 (2018).
13. G. H. Cartledge, *J. Am. Chem. Soc.* **50**, 2855–2863 (1928).
14. C. Delmas, J.-J. Braconnier, C. Fouassier, P. Hagenmuller, *Solid State Ion.* **3-4**, 165–169 (1981).
15. Y.-J. Shin, M.-Y. Yi, *Solid State Ion.* **132**, 131–141 (2000).
16. G. Singh *et al.*, *Chem. Mater.* **28**, 5087–5094 (2016).
17. C. Delmas, C. Fouassier, P. Hagenmuller, *Mater. Res. Bull.* **11**, 1483–1488 (1976).
18. J. Rouxel, *J. Solid State Chem.* **17**, 223–229 (1976).
19. D. Kim, M. Cho, K. Cho, *Adv. Mater.* **29**, 1701788 (2017).
20. T. Sato, K. Sato, W. Zhao, Y. Kajiya, N. Yabuuchi, *J. Mater. Chem. A Mater. Energy Sustain.* **6**, 13943–13951 (2018).
21. T. Uyama, K. Mukai, I. Yamada, *Inorg. Chem.* **58**, 6684–6695 (2019).
22. M. H. Han *et al.*, *Electrochim. Acta* **182**, 1029–1036 (2015).
23. M. Bianchini *et al.*, *Nat. Mater.* **19**, 1088–1095 (2020).

ACKNOWLEDGMENTS

Funding: This work was supported by the National Natural Science Foundation of China (51725206, 51421002, 21773303), National Key Technologies R&D Program of China (2016YFB0901500), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA21070500), Youth Innovation Promotion Association, Chinese Academy of Sciences (2020006), Beijing Municipal Science and Technology Commission (Z181100004718008), Beijing Natural Science Fund-Haidian Original Innovation Joint Fund (L182056), and the Netherlands Organization for Scientific Research (NWO) (under the VICI grant no. 16122). Computations were performed on the Niagara supercomputer at the SciNet HPC Consortium. SciNet is funded by the Canada Foundation for Innovation; the Government of Ontario, Ontario Research Fund-Research Excellence, and the University of Toronto. C.Z. also thanks to the State Scholarship Fund of China Scholarship Council (CSC). **Author contributions:** Y.-S.H. conceived this research and supervised this work with M.W., C.Z., and Q.W., who conceptualized the ionic potential method and developed the calculation on examples of Na-/Li-/K-ion layered oxides. C.Z. and Q.W. performed synthesis procedures, experimental investigation of $\text{NaLi}_{1/3}\text{Ti}_{1/6}\text{Mn}_{1/2}\text{O}_2$ and $\text{Na}_{5/6}\text{Li}_{5/18}\text{Mn}_{13/18}\text{O}_2$ materials, and software programming to process and present collected data. F.D. synthesized the Na-Li-Cu-Fe-Mn-O materials. Z.Y., B.S.-L., and A.A.-G. predict Na-ion layered oxides tested by cationic potential. J.W. and X.B. performed STEM observation and analysis. C.Z., Q.W., Z.Y., M.W., Y.L., C.D., and Y.-S.H. wrote the manuscript. All authors participated in analyzing the experimental results and preparing the manuscript. C.Z., Q.W., and Z.Y. contributed equally to this work. **Competing interests:** All authors declare that they have no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6517/708/suppl/DC1
Materials and Methods
Supplementary Text
Figs. S1 to S15
Tables S1 to S19
References (24–169)

4 August 2019; accepted 18 September 2020
10.1126/science.aay9972

ECOLOGICAL MONITORING

Ecological insights from three decades of animal movement tracking across a changing Arctic

Sarah C. Davidson^{1,2,3}, Gil Bohrer^{1*}, Eliezer Gurarie^{4,5}, Scott LaPoint^{2,6,7}, Peter J. Mahoney⁸, Natalie T. Boelman⁷, Jan U. H. Eitel⁹, Laura R. Prugh⁸, Lee A. Vierling⁹, Jyoti Jennewein⁹, Emma Grier⁴, Ophélie Couriot^{4,10}, Alicia P. Kelly¹¹, Arjan J. H. Meddens¹², Ruth Y. Oliver^{7,13,14}, Roland Kays¹⁵, Martin Wikelski^{2,3}, Tomas Aarvak¹⁶, Joshua T. Ackerman¹⁷, José A. Alves^{18,19}, Erin Bayne²⁰, Bryan Bedrosian²¹, Jerrold L. Belant²², Andrew M. Berdahl²³, Alicia M. Berlin²⁴, Dominique Berteaux²⁵, Joël Bêty²⁵, Dmitrijs Boiko^{26,27,28}, Travis L. Booms²⁹, Bridget L. Borg³⁰, Stan Boutin²⁰, W. Sean Boyd³¹, Kane Brides³², Stephen Brown³³, Victor N. Bulyuk³⁴, Kurt K. Burnham³⁵, David Cabot³⁶, Michael Casazza¹⁷, Katherine Christie³⁷, Erica H. Craig³⁸, Shanti E. Davis³⁹, Tracy Davison⁴⁰, Dominic Demma⁴¹, Christopher R. DeSorbo⁴², Andrew Dixon⁴³, Robert Domenech⁴⁴, Götz Eichhorn^{45,46}, Kyle Elliott⁴⁷, Joseph R. Evenson⁴⁸, Klaus-Michael Exo⁴⁹, Steven H. Ferguson⁵⁰, Wolfgang Fiedler^{2,3}, Aaron Fisk⁵¹, Jérôme Fort⁵², Alastair Franke^{20,53}, Mark R. Fuller⁵⁴, Stefan Garthe⁵⁵, Gilles Gauthier⁵⁶, Grant Gilchrist⁵⁷, Petr Glazov⁵⁸, Carrie E. Gray⁵⁹, David Grémillet^{60,61}, Larry Griffin³², Michael T. Hallworth^{62,63}, Autumn-Lynn Harrison⁶², Holly L. Hennin^{31,64}, J. Mark Hipfner⁶⁵, James Hodson⁶⁶, James A. Johnson⁶⁷, Kyle Joly⁶⁸, Kimberly Jones⁴¹, Todd E. Katzner⁶⁹, Jeff W. Kidd⁷⁰, Elly C. Knight²⁰, Michael N. Kochert⁶⁹, Andrea Kölzsch^{2,3,71}, Helmut Kruckenberg⁷¹, Benjamin J. Lagassé⁷², Sandra Lai²⁵, Jean-François Lamarre⁷³, Richard B. Lanctot⁶⁷, Nicholas C. Larter⁷⁴, A. David M. Latham^{20,75}, Christopher J. Latty⁷⁶, James P. Lawler⁷⁷, Don-Jean Léandri-Breton²⁵, Hansoo Lee⁷⁸, Stephen B. Lewis⁷⁹, Oliver P. Love⁶⁴, Jesper Madsen⁸⁰, Mark Maffei³⁹, Mark L. Mallory⁸¹, Buck Mangipane⁸², Mikhail Y. Markovets³⁴, Peter P. Marra⁸³, Rebecca McGuire⁸⁴, Carol L. McIntyre³⁰, Emily A. McKinnon⁸⁵, Tricia A. Miller^{86,87}, Sander Moonen⁴⁹, Tong Mu⁸⁸, Gerhard J. D. M. Müskens⁸⁹, Janet Ng²⁰, Kerry L. Nicholson²⁹, Ingar Jostein Øien¹⁶, Cory Overton¹⁷, Patricia A. Owen³⁰, Allison Patterson⁴⁷, Aevor Petersen⁹⁰, Ivan Pokrovsky^{2,91,92}, Luke L. Powell^{62,93,94}, Rui Prieto⁹⁵, Petra Quillfeldt⁹⁶, Jennie Rausch⁹⁷, Kelsey Russell⁹⁸, Sarah T. Saalfeld⁶⁷, Hans Schekkerman⁹⁹, Joel A. Schmutz¹⁰⁰, Philipp Schwemmer⁵⁵, Dale R. Seip¹⁰¹, Adam Shreading⁴⁴, Mónica A. Silva^{95,102}, Brian W. Smith¹⁰³, Fletcher Smith^{104,105}, Jeff P. Smith^{106,107}, Katherine R. S. Snell^{2,108}, Aleksandr Sokolov⁹², Vasilij Sokolov¹⁰⁹, Diana V. Solovyeva⁹¹, Mathew S. Sorum¹¹⁰, Grigori Tertitski⁵⁸, J. F. Therrien^{56,111}, Kasper Thorup¹⁰⁸, T. Lee Tibbitts¹⁰⁰, Ingrid Tulp¹¹², Brian D. Uher-Koch¹⁰⁰, Rob S. A. van Bemmelen^{112,113}, Steven Van Wilgenburg¹¹⁴, Andrew L. Von Duyke¹¹⁵, Jesse L. Watson²⁰, Bryan D. Watts¹⁰⁴, Judy A. Williams⁶⁶, Matthew T. Wilson⁴⁸, James R. Wright¹¹⁶, Michael A. Yates¹¹⁷, David J. Yurkowski^{50,85}, Ramūnas Žydelis¹¹⁸, Mark Hebblewhite⁵

The Arctic is entering a new ecological state, with alarming consequences for humanity. Animal-borne sensors offer a window into these changes. Although substantial animal tracking data from the Arctic and subarctic exist, most are difficult to discover and access. Here, we present the new Arctic Animal Movement Archive (AAMA), a growing collection of more than 200 standardized terrestrial and marine animal tracking studies from 1991 to the present. The AAMA supports public data discovery, preserves fundamental baseline data for the future, and facilitates efficient, collaborative data analysis. With AAMA-based case studies, we document climatic influences on the migration phenology of eagles, geographic differences in the adaptive response of caribou reproductive phenology to climate change, and species-specific changes in terrestrial mammal movement rates in response to increasing temperature.

The Arctic and adjacent regions are experiencing the most rapid climate and environmental changes on Earth, caused primarily by anthropogenic greenhouse gas emissions (1). Notable trends include warming winter temperatures, ice loss, and earlier spring snowmelt. These changes profoundly affect conditions experienced by animals, including food availability, interspecific

competition, predation, and increased human disturbances (2). Impacts of climate change on Arctic vertebrates include rapid poleward range shifts (3, 4); phenological trophic mismatches (5); and changes in migration (6), foraging, and predator–prey dynamics (7). Because rapid environmental change in the Arctic challenges the ability of the region's fauna to adapt, a primary response will likely occur through phenotypic plasticity in the patterns, locations, and timing of their movements

(2). Documenting and understanding these changes requires multidecadal, pan-Arctic data at multiple trophic levels.

We demonstrate the ecological utility of the Arctic Animal Movement Archive (AAMA), an active, collaborative collection of animal tracking datasets (supplementary materials). Marine ecology archives, such as IOOS-ATN, IMOS, OBIS-SEAMAP, and RAATD (8), provide insight regarding space use, movement, and connectivity (9–11). Terrestrial animal movement archives are rare and tend to have a regional or taxonomic focus (12). AAMA is the first Arctic-focused archive with both terrestrial and marine data and is hosted on the global Movebank database. The geographic scope of the AAMA (Fig. 1) includes the Arctic, Arctic marine, and subarctic “boreal forests/taiga” regions defined elsewhere (13, 14) (see also supplementary materials). Currently, the archive contains more than 15,000,000 occurrences of 8000 individuals representing 86 species, from 1991 to the present (figs. S1 and S2 and tables S1 to S4). Combining data from multiple AAMA studies, we show evidence of (i) climate drivers of golden eagle migration phenology, (ii) climate adaptation of parturition by caribou, and (iii) consequences of increased temperature and precipitation on movements of mammalian predators and herbivores.

Behavioral flexibility enables migrants to optimize energy expenditure during migration and adjust arrival at summering grounds (15, 16). We used tracking data from 103 individuals during 1993 to 2017 [supplementary materials (case study 1) and table S5] to examine arrival timing to breeding grounds of northward-migrating golden eagles (“summering”), modeling it with predictors for age, sex, summering onset latitude, year, and the preceding winter's mean Pacific decadal oscillation index (PDO).

Mean summering date changed slowly over 25 years (−0.5 days/year). The long-term trend differed among age classes, with adults arriving earliest, then subadults, and then juveniles, and it was influenced by winter climate (PDO) (Fig. 2 and tables S8 and S9). Eagles of all age classes began summering later at northern latitudes (1.08 days/degree). The significant interaction of year and previous “warm-phase” PDO explains earlier summering dates for subadults and juveniles, highlighting their known responsiveness to environmental conditions (16). These warm-phase winters cause a warmer and drier climate with reduced snowpack and an earlier snow-free date. Earlier adult arrival to summering grounds should result from selection and competition for territories, yet local climatic variables affect eagle condition before, and energy expenditure during, northward migration (16). For subadults sampled after 2011, the direct effect of PDO is significant (−8.27 days), whereas the full subadult dataset does not show a

Author affiliations are listed at the end of this paper.

*Corresponding author. Email: bohrer.17@osu.edu

significant effect of winter PDO (Fig. 2). This period-related difference in inference of climatic drivers highlights the importance of compiling long-term, multigenerational observations. Given the importance of the winter PDO and known impacts of global climate change, golden eagles could face age-specific challenges during migration and at their warming Arctic summering grounds.

The timing of parturition is a key to the demography of wildlife populations and can be an adaptive response to climate shifts (17). For many mammals, the period from late pregnancy through weaning has the highest energetic demands and thus is timed to occur when vegetation productivity is highest (18). Caribou occur in five different ecotypes (Fig. 3) across boreal and Arctic North America and are facing global declines (19). On the basis of data from 917 individuals during 2000 to 2017 in northern Canada, we used characteristic patterns of low movement during the calving season to estimate 1630 parturition dates in five populations of barren-ground, northern and southern boreal woodland, and northern and southern mountain woodland caribou [supplementary materials (case study 2) and table S6].

We found differences in parturition timing and trends among the five populations. The southern and northern boreal populations calved earliest, followed by northern and southern mountain populations (table S10). Barren-ground caribou calved later despite occupying a similar latitudinal range as the northern boreal caribou (Fig. 3). Most importantly, barren-ground and northern woodland caribou, but not southern woodland caribou, exhibited significant trends toward earlier parturition [0.4 to 1.1 days/year (table S10)]. This is the first continental-scale retrospective evidence of potential adaptive responses to climate trends by caribou.

Animals conserve energy by modifying their behavior in response to weather conditions, with important implications for individual fitness and species resilience under climate change (20). We tested for effects of temperature and precipitation on seasonal movement rates (in meters per minute) using records from 1720 individuals of two herbivore and three predator species (black bear, grizzly bear, caribou, moose, and wolf) during 1998 to 2019 [supplementary materials (case study 3) and table S7]. We predicted that winter movement rates would decline relative to summer, when energetic costs of self-maintenance would be highest. Rate would also decline within seasons, during weather conditions that increase the energetic cost of movement (e.g., snow that increases energy requirements for movement or higher ambient temperatures during the summer that accelerate metabolism).

All species exhibited lower movement rates during winter relative to summer (Fig. 4). As

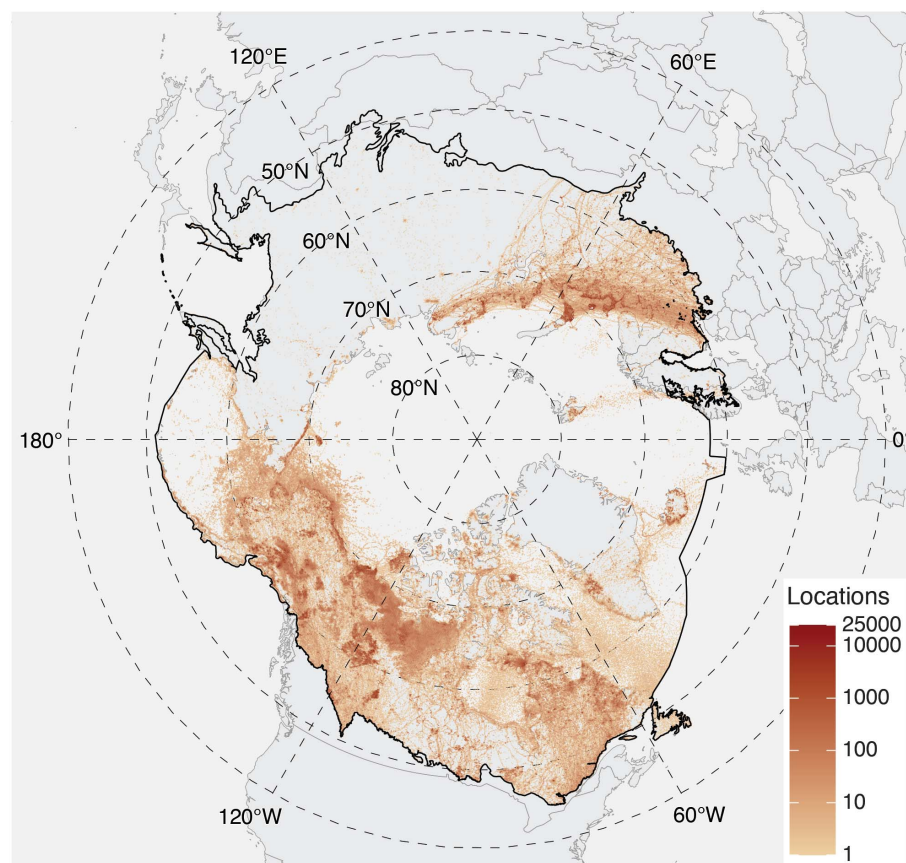


Fig. 1. Map of the AAMA boundary and data. Density of animal locations (number of observations per ~100 km²) at logarithmic scale characterizes data availability, not animal density or utilization.

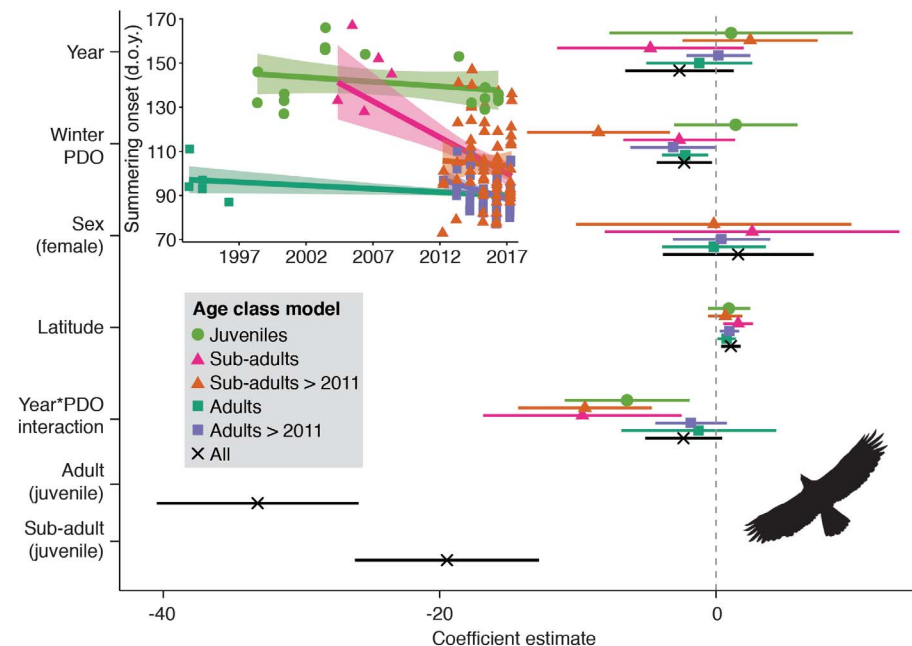


Fig. 2. Changes in the onset date of golden eagles' summering. Coefficient estimates (±95% confidence intervals) reflecting age-specific changes in response to year, previous winter PDO, sex (reference: females), latitude, interaction of year and PDO, and age class [reference: juveniles (tables S8 and S9)]. (Inset) Time series of model-estimated summering. d.o.y., day of year.

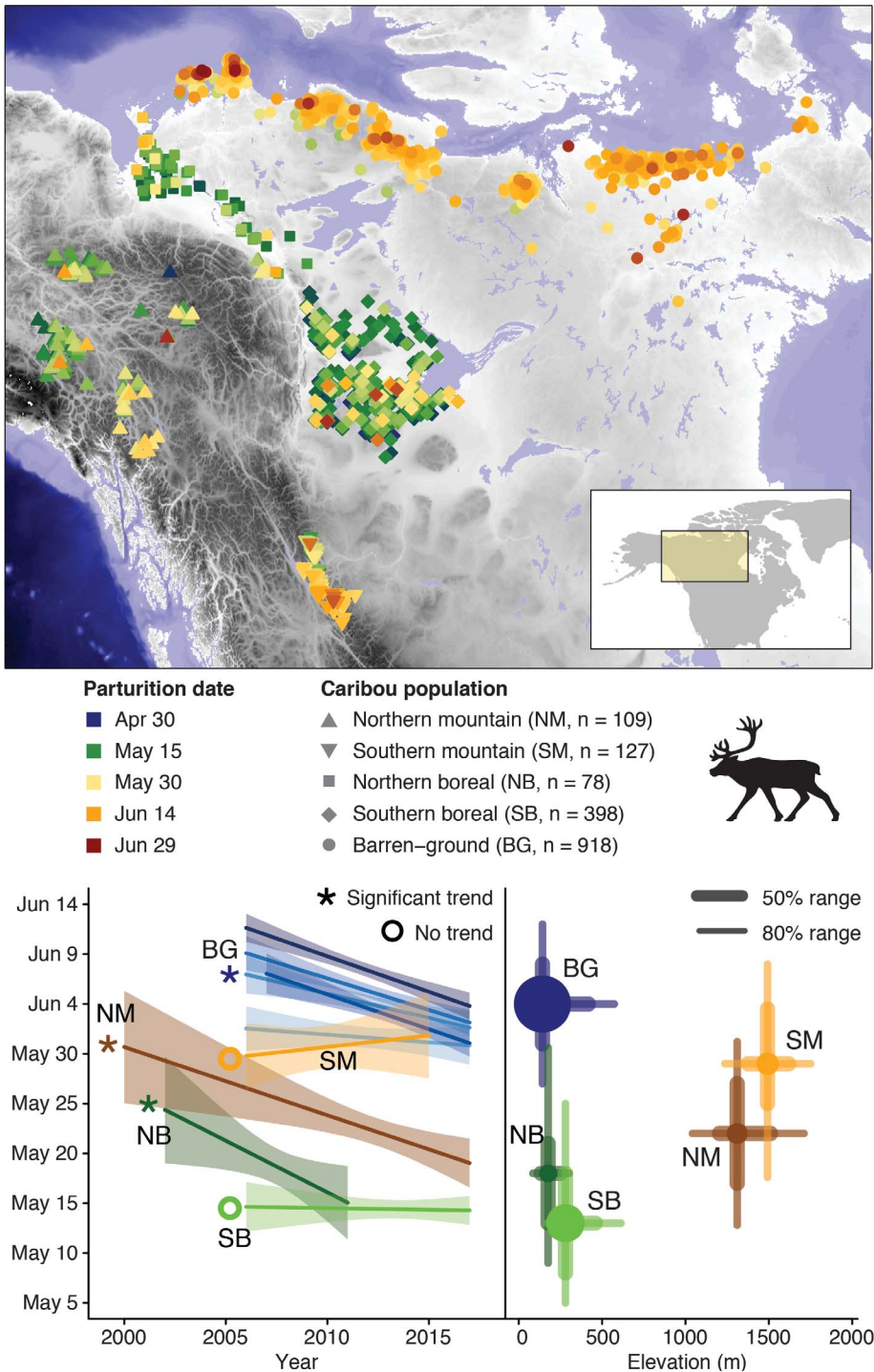


Fig. 3. Climate change adaptation of parturition times (PT) of caribou. (Top) PT by population. (Bottom left) PT trends by population, including five barren-ground subpopulations. (Bottom right) PT dates by elevation.

temperatures increased in summer, wolves and black bears slowed their movement rates, whereas moose increased their movement rates. In winter, only barren-ground caribou increased movement rates as temperature increased. Snow impeded wolves, boreal caribou, and moose,

whereas all species were generally insensitive to summer precipitation. These patterns may reflect asynchronous responses to climate change within and across trophic levels. Climate-driven variation in animal activity is likely to affect species interactions, altering energy

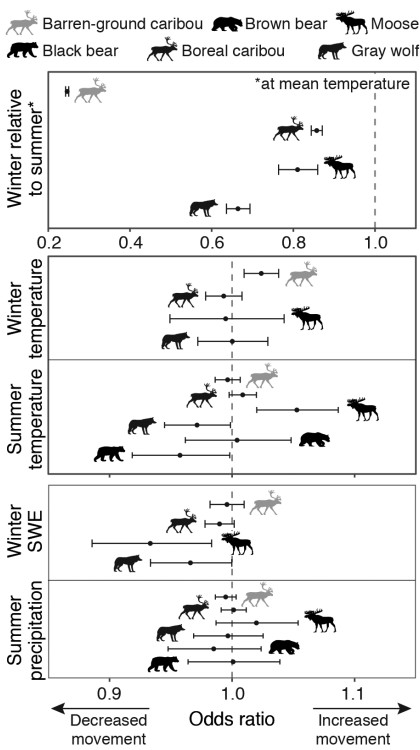


Fig. 4. Changes in species-specific movement rates in response to daily maximum temperature, summer precipitation, and winter snow-water equivalent (SWE). Odds ratios for continuous covariates represent the positive or negative change in movement rates per one unit change in temperature or precipitation, respectively. Ratios were identified as neutral if credible intervals overlapped with 1.0.

expenditure, encounter rates, and foraging success with demographic implications for both predators and prey.

As we demonstrate, the AAMA provides a solution to Arctic data collection and sharing challenges. It serves as a critical baseline and resource to identify early signals of local or large-scale changes in animal distribution, movement responses, and adaptive traits. Continued shifts in phenology in the Arctic pose challenges to migratory species that encounter changing seasonal fluctuations along migration routes and at Arctic summering and southern wintering grounds (21). Key drivers of population responses, such as migration, parturition, and foraging movement, are undergoing rapid changes, suggesting that climate change is affecting animals in ways that will shape the future of the Arctic.

REFERENCES AND NOTES

1. IPCC, "Climate change 2014: Synthesis report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change" (IPCC, 2015).

2. O. Gilg *et al.*, *Ann. N. Y. Acad. Sci.* **1249**, 166–190 (2012).
3. M. Fossheim *et al.*, *Nat. Clim. Chang.* **5**, 673–677 (2015).
4. I.-C. Chen, J. K. Hill, R. Ohlemüller, D. B. Roy, C. D. Thomas, *Science* **333**, 1024–1026 (2011).
5. S. T. Saalfeld, R. B. Lanctot, *Ecol. Evol.* **7**, 10492–10502 (2017).
6. D. H. Ward *et al.*, *J. Avian Biol.* **47**, 197–207 (2016).
7. R. F. Rockwell, L. J. Gormezano, D. N. Koons, *Oikos* **120**, 696–709 (2011).
8. Y. Ropert-Coudert *et al.*, *Sci. Data* **7**, 94 (2020).
9. M. A. Hindell *et al.*, *Nature* **580**, 87–92 (2020).
10. G. C. Hays *et al.*, *Trends Ecol. Evol.* **34**, 459–473 (2019).
11. S. Brodie *et al.*, *Sci. Rep.* **8**, 3717 (2018).
12. F. Cagnacci *et al.*, *Oikos* **120**, 1790–1803 (2011).
13. J. I. Murray, L. Hacquebord, D. J. Gregor, H. Loeng, Eds., in “AMAP assessment report: Arctic pollution issues” (Arctic Monitoring and Assessment Programme, 1998), chap. 2, pp. 9–23.
14. The Nature Conservancy, *tnc_terr_ecoregions* (2009); <http://maps.tnc.org/metadata/TerrEcos.xml>.
15. D. W. Winkler *et al.*, *Mov. Ecol.* **2**, 10 (2014).
16. T. A. Miller *et al.*, *Ibis* **158**, 116–134 (2016).
17. T. Bonnet *et al.*, *PLOS Biol.* **17**, e3000493 (2019).
18. D. C. Stoner, J. O. Sexton, J. Nagol, H. H. Bernalles, T. C. J. Edwards Jr., *PLOS ONE* **11**, e0148780 (2016).
19. L. S. Vors, M. S. Boyce, *Glob. Change Biol.* **15**, 2626–2633 (2009).
20. A. Clarke, K. P. P. Fraser, *Funct. Ecol.* **18**, 243–251 (2004).
21. J. A. Gilg *et al.*, *Proc. Biol. Sci.* **281**, 20132161 (2013).
22. G. Bohrer *et al.*, Data from “Ecological insights from three decades of animal movement tracking across a changing Arctic.” Dryad (2020); <https://doi.org/10.5061/dryad.k98sf7m4m>.
- ⁸School of Environmental and Forest Sciences, University of Washington, Seattle, WA, USA. ⁹Department of Natural Resources and Society, University of Idaho, Moscow, ID, USA. ¹⁰National Socio-Environmental Synthesis Center, Annapolis, MD, USA. ¹¹Department of Environment and Natural Resources, Government of the Northwest Territories, Fort Smith, NT, Canada. ¹²School of the Environment, Washington State University, Pullman, WA, USA. ¹³Department of Ecology and Evolutionary Biology, Yale University, New Haven, CT, USA. ¹⁴Center for Biodiversity and Global Change, Yale University, New Haven, CT, USA. ¹⁵College of Natural Resources, North Carolina State University, Raleigh, NC, USA. ¹⁶BirdLife Norway, Trondheim, Norway. ¹⁷U.S. Geological Survey, Western Ecological Research Center, Dixon Field Station, Dixon, CA, USA. ¹⁸Department of Biology & CESAM, University of Aveiro, Aveiro, Portugal. ¹⁹South Iceland Research Centre, University of Iceland, Laugarvatn, Iceland. ²⁰Department of Biological Sciences, University of Alberta, Edmonton, AB, Canada. ²¹Teton Raptor Center, Jackson Hole, WY, USA. ²²Global Wildlife Conservation Center, College of Environmental Science and Forestry, State University of New York, Syracuse, NY, USA. ²³School of Aquatic & Fishery Sciences, University of Washington, Seattle, WA, USA. ²⁴U.S. Geological Survey, Patuxent Wildlife Research Center, Laurel, MD, USA. ²⁵Centre d'études nordiques, Université du Québec à Rimouski, Rimouski, QC, Canada. ²⁶Latvian National Museum of Natural History, Riga, Latvia. ²⁷Institute of Biology, University of Latvia, Salaspils, Latvia. ²⁸Latvian Swan Research Society, Kalnciems, Latvia. ²⁹Alaska Department of Fish and Game, Fairbanks, AK, USA. ³⁰National Park Service, Denali National Park and Preserve, Denali Park, AK, USA. ³¹Science & Technology Branch, Environment & Climate Change Canada, Delta, BC, Canada. ³²Wildfowl & Wetlands Trust, Slimbridge, UK. ³³Manomet, Inc., Saxtons River, VT, USA. ³⁴Biological Station Rybachy, Zoological Institute of Russian Academy of Sciences, St. Petersburg, Russia. ³⁵High Arctic Institute, Orion, IL, USA. ³⁶School of Biological, Earth and Environmental Sciences, University College Cork, Cork, Ireland. ³⁷Alaska Department of Fish and Game, Anchorage, AK, USA. ³⁸Aquila Environmental, Fairbanks, AK, USA. ³⁹High Arctic Gull Research Group, Bamfield, BC, Canada. ⁴⁰Department of Environment and Natural Resources, Government of the Northwest Territories, Inuvik, NT, Canada. ⁴¹Alaska Department of Fish and Game, Palmer, AK, USA. ⁴²Biodiversity Research Institute, Portland, ME, USA. ⁴³Reneco International Wildlife Consultants, Abu Dhabi, United Arab Emirates. ⁴⁴Raptor View Research Institute, Missoula, MT, USA. ⁴⁵Vogeltrekstation—Dutch Centre for Avian Migration and Demography, Wageningen, Netherlands. ⁴⁶Department of Animal Ecology, Netherlands Institute of Ecology (NIOO-KNAW), Wageningen, Netherlands. ⁴⁷Department of Natural Resource Sciences, McGill University, Ste Anne-de-Bellevue, QC, Canada. ⁴⁸Washington Department of Fish and Wildlife, Olympia, WA, USA. ⁴⁹Institute for Avian Research “Vogelwarte Helgoland,” Wilhelmshaven, Germany. ⁵⁰Fisheries and Oceans Canada, Winnipeg, MB, Canada. ⁵¹Great Lakes Institute for Environmental Research, School of the Environment, University of Windsor, Windsor, ON, Canada. ⁵²Littoral Environnement et Sociétés (LIENSs), CNRS, La Rochelle University, La Rochelle, France. ⁵³Arctic Raptor Project, Rankin Inlet, NU, Canada. ⁵⁴Boise State University, Raptor Research Center, Boise, ID, USA. ⁵⁵Research and Technology Centre (FTZ), Kiel University, Büsum, Germany. ⁵⁶Département de Biologie & Centre d'Études Nordiques, Université Laval, Quebec City, QC, Canada. ⁵⁷Environment & Climate Change Canada, National Wildlife Research Centre, Carleton University, Ottawa, ON, Canada. ⁵⁸Institute of Geography, Russian Academy of Sciences, Moscow, Russia. ⁵⁹School of Biology and Ecology, University of Maine, Orono, ME, USA. ⁶⁰Centre d'Études Biologiques de Chizé, CNRS, La Rochelle University, Villiers en Bois, France. ⁶¹Percy Fitzpatrick Institute of African Ornithology, University of Cape Town, Rondebosch, South Africa. ⁶²Migratory Bird Center, Smithsonian Conservation Biology Institute, National Zoological Park, Washington DC, USA. ⁶³Northeast Climate Adaptation Science Center, University of Massachusetts Amherst, Amherst, MA, USA. ⁶⁴Department of Integrative Biology, University of Windsor, Windsor, ON, Canada. ⁶⁵Environment & Climate Change Canada, Pacific Wildlife Research Centre, Delta, BC, Canada. ⁶⁶Department of Environment and Natural Resources, Government of the Northwest Territories, Yellowknife, NT, Canada. ⁶⁷U.S. Fish & Wildlife Service, Migratory Bird Management, Anchorage, AK, USA. ⁶⁸National Park Service, Gates of the Arctic National Park & Preserve, Fairbanks, AK, USA. ⁶⁹U.S. Geological Survey, Forest and Rangeland Ecosystem Science Center, Boise, ID, USA. ⁷⁰Kidd Biological, Inc., Anacortes, WA, USA. ⁷¹Institute for Wetlands and Waterbird Research e.V., Verden (Aller), Germany. ⁷²Department of Integrative Biology, University of Colorado, Denver, CO, USA. ⁷³Polar Knowledge Canada, Cambridge Bay, NU, Canada. ⁷⁴Department of Environment and Natural Resources, Government of the Northwest Territories, Fort Simpson, NT, Canada. ⁷⁵Manaaki Whenua—Landcare Research, Lincoln, New Zealand. ⁷⁶U.S. Fish & Wildlife Service, Arctic National Wildlife Refuge, Fairbanks, AK, USA. ⁷⁷National Park Service, Alaska Inventory and Monitoring Program, Anchorage, AK, USA. ⁷⁸Korea Institute of Environmental Technology, Yuseonggu, Daejeon, Republic of Korea. ⁷⁹U.S. Fish & Wildlife Service, Juneau, AK, USA. ⁸⁰Department of Bioscience—Kalø, Aarhus University, Rønde, Denmark. ⁸¹Biology Department, Acadia University, Wolfville, NS, Canada. ⁸²National Park Service, Lake Clark National Park and Preserve, Anchorage, AK, USA. ⁸³Department of Biology and the McCourt School of Public Policy, Georgetown University, Washington, DC, USA. ⁸⁴Wildlife Conservation Society, Arctic Beringia Program, Fairbanks, AK, USA. ⁸⁵University of Manitoba, Winnipeg, MB, Canada. ⁸⁶Conservation Science Global, Inc., West Cape May, NJ, USA. ⁸⁷Division of Forestry and Natural Resources, West Virginia University, Morgantown, WV, USA. ⁸⁸Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ, USA. ⁸⁹Wageningen Environmental Research, Wageningen University & Research, Wageningen, Netherlands. ⁹⁰Independent researcher, Reykjavik, Iceland. ⁹¹Laboratory of Ornithology, Institute of Biological Problems of the North FEB RAS, Magadan, Russia. ⁹²Arctic Research Station of Institute of Plant and Animal Ecology UB, RAS, Labytnangi, Yamal-Nenets Autonomous District, Russia. ⁹³Durham University, Durham, UK. ⁹⁴University of Glasgow, Glasgow, Scotland. ⁹⁵Marine and Environmental Sciences Centre, Institute of Marine Research and Okeanos R&D Centre, University of the Azores, Horta, Portugal. ⁹⁶Justus-Liebig University, Gießen, Germany. ⁹⁷Environment & Climate Change Canada, Yellowknife, NT, Canada. ⁹⁸Environment Yukon, Whitehorse, YT, Canada. ⁹⁹SOVON, Nijmegen, Netherlands. ¹⁰⁰U.S. Geological Survey Alaska Science Center, Anchorage, AK, USA. ¹⁰¹British Columbia Ministry of Environment, Prince George, BC, Canada. ¹⁰²Biology Department, Woods Hole Oceanographic Institution, Woods Hole, MA, USA. ¹⁰³U.S. Fish & Wildlife Service, Migratory Bird Management, Denver, CO, USA. ¹⁰⁴Center for Conservation Biology, College of William & Mary, Williamsburg, VA, USA. ¹⁰⁵Georgia Department of Natural Resources, Brunswick, GA, USA. ¹⁰⁶HawkWatch International, Salt Lake City, UT, USA. ¹⁰⁷H. T. Harvey & Associates, Los Gatos, CA, USA. ¹⁰⁸Center for Macroecology, Evolution and Climate, Globe Institute, University of Copenhagen, Copenhagen, Denmark. ¹⁰⁹Institute of Plant and Animal Ecology, Ural Division Russian Academy of Sciences, Ekaterinburg, Russia. ¹¹⁰National Park Service, Yukon-Charley Rivers National Preserve, Central Alaska Inventory and Monitoring Network, Fairbanks, AK, USA. ¹¹¹Hawk Mountain Sanctuary, Kempton, PA, USA. ¹¹²Wageningen Marine Research, IJmuiden, Netherlands. ¹¹³Bureau Waardenburg, Culemborg, Netherlands. ¹¹⁴Canadian Wildlife Service, Environment & Climate Change Canada, Saskatoon, SK, Canada. ¹¹⁵North Slope Borough, Department of Wildlife Management, Utqiagvik, AK, USA. ¹¹⁶School of Environment and Natural Resources, The Ohio State University, Columbus, OH, USA. ¹¹⁷Earthspan Foundation, Minden, NV, USA. ¹¹⁸Ornitela UAB, Vilnius, Lithuania.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6517/712/suppl/DC1
Materials and Methods
Figs. S1 and S2
Tables S1 to S10
References (23–64)
MDAR Reproducibility Checklist

12 March 2020; resubmitted 16 March 2020
Accepted 15 September 2020
10.1126/science.abb7080

PALEOCLIMATE

Phasing of millennial-scale climate variability in the Pacific and Atlantic Oceans

Maureen H. Walczak^{1,2*}, Alan C. Mix¹, Ellen A. Cowan³, Stewart Fallon², L. Keith Fifield², Jay R. Alder^{1,4†}, Jianghui Du^{1,5†}, Brian Haley^{1†}, Tim Hobern^{2†}, June Padman^{1†}, Summer K. Praetorius^{6†}, Andreas Schmittner^{1†}, Joseph S. Stoner^{1†}, Sarah D. Zellers^{7†}

New radiocarbon and sedimentological results from the Gulf of Alaska document recurrent millennial-scale episodes of reorganized Pacific Ocean ventilation synchronous with rapid Cordilleran Ice Sheet discharge, indicating close coupling of ice-ocean dynamics spanning the past 42,000 years. Ventilation of the intermediate-depth North Pacific tracks strength of the Asian monsoon, supporting a role for moisture and heat transport from low latitudes in North Pacific paleoclimate. Changes in carbon-14 age of intermediate waters are in phase with peaks in Cordilleran ice-rafted debris delivery, and both consistently precede ice discharge events from the Laurentide Ice Sheet, known as Heinrich events. This timing precludes an Atlantic trigger for Cordilleran Ice Sheet retreat and instead implicates the Pacific as an early part of a cascade of dynamic climate events with global impact.

During the last glacial period, the Laurentide Ice Sheet of North America experienced recurrent, unstable millennial-scale retreat events, characterized by episodic iceberg discharge inferred from the presence of apparently ice-rafted detrital sedimentary layers far into the Atlantic Ocean (1). These so-called Heinrich events are among the most abrupt climate perturbations of Earth's recent past (2). Various triggering mechanisms have been theorized, including reduction of the Atlantic Meridional Overturning Circulation (AMOC) (3) warming of subsurface waters (4–6), internal dynamics of the large Laurentide Ice Sheet (7), and/or sea-level rise triggered by episodic failure of another (presumably European) ice sheet (8). Regardless of origin, changes in oceanic and atmospheric circulation associated with the Heinrich events correspond to global perturbations, including weak intervals of the Asian monsoon (9), and antiphased warming in Antarctica (10).

Marine sedimentary records spanning the Northeast Pacific document periods of high ice discharge analogous to the Heinrich events off the west coast of North America, but with uncertain timing and cause (11–14). Here, on the basis of a detailed radiocarbon chronology in the Gulf of Alaska, we assess the timing of the Pacific ice discharge events from the Cordilleran Ice Sheet and changes in apparent sub-

surface ocean radiocarbon ages and establish phasing relationships between North Pacific and North Atlantic events. We find that Pacific discharge events and intermediate water ventilation changes precede events in the North Atlantic and are early parts of a dynamic cascade of global climate changes, including Antarctic warming and atmospheric CO₂ rise. We thus reject hypotheses that these North Pacific millennial-scale climate events originate as passive teleconnected responses to the Laurentide/North Atlantic or Antarctic.

During the last glacial period, the Alaskan margin hosted ice streams draining the northwestern limb of the Cordilleran Ice Sheet, evidenced by a series of sedimented troughs crossing the continental shelf (15). Although smaller than the Laurentide Ice Sheet, the Cordilleran Ice Sheet at its Last Glacial Maximum (LGM) extent was slightly larger than the modern-day Greenland Ice Sheet (16). IODP Site U1419 (59°31.9'N, 144°8.0'W, 690-m water depth; Fig. 1) was drilled on the Gulf of Alaska slope in ~690-m water depth on the continental slope ~75 km seaward of the Bering Glacier, a modern remnant of a major ice stream that may have routed ~15% of drainage from the Cordilleran Ice Sheet (fig. S1). Four drill holes yielded a continuous stratigraphic splice extending back >90 m adjusted core composite depth below seafloor, CCSF-B (17).

A Bayesian age model based on 250 ¹⁴C measurements of planktic and benthic foraminifera, and limited correlation points beyond the range of ¹⁴C (table S1 and fig. S2), indicates that the base of the stratigraphic splice was deposited ~55,000 years before present (yr B.P.) (tables S3 and S4 and Fig. 2). The chronology discussed here is based on calibration of radiocarbon dates using the Marine13 curve (18);

however, our conclusions are insensitive to calibration on the Marine20 curve (19), and that alternate age model is presented and discussed in the supplementary materials. Corrected ages and sediment accumulation rates, and their uncertainties, are resolved in 500-year intervals to ~60,000 yr B.P. (tables S4 and S5). Accumulation rates range from as low as ~10 cm per thousand years (kyr) in mid-Holocene time (average 50 ± 30 cm kyr⁻¹, 0 to 11,700 BP) to peaks of ~800 cm kyr⁻¹ during ice discharge events of the latest Pleistocene (average 200 ± 160 cm kyr⁻¹, 12,000 to 50,000 BP; fig. S3).

Each interval of anomalously high sediment accumulation rate off southeast Alaska contains a high concentration of coarse (sand-size or greater) grains not associated with turbidites, interpreted as ice-rafted detritus (IRD) (see materials and methods). Mean IRD mass accumulation rates (MARs) range from zero (indicating that glaciers retreated from the ocean either onto land or behind fjord sills sufficiently shallow to inhibit iceberg fluxes) to as high as 40 g cm⁻² kyr⁻¹. We name these episodes of high IRD MAR “Siku events” (the Iñupiat/Inuit word for ice, abbreviated here as “S events”), which we define operationally as IRD MAR >12 g cm⁻² kyr⁻¹ averaged over a time span of 500 years or longer. This definition yields events S1 (peak at 17.0 to 18.0 kyr B.P.), S2 (25.0 to 27.0 kyr B.P.), S3 (29.5 to 30.5 kyr B.P.), S4 (39.5 to 42.0 kyr B.P.), and perhaps S5 (~54 to 56 kyr B.P., although the age model and IRD MAR is imprecise in this interval) (Fig. 3 and fig. S9). The Siku events represent a massive influx of icebergs from regional retreat of marine-terminating outlet glaciers from the Cordilleran Ice Sheet. This association is analogous to that of North Atlantic Heinrich events (20) and is consistent with evidence for anomalously high sediment fluxes associated with recent tidewater glacier retreat in Alaska (21).

The lithology of IRD grains at Site U1419 varies widely, reflecting both proximal sources in the Chugach and St. Elias Ranges drained by the Bering Glacier, as well as more distal sources, including the Alexander terrane of southeast Alaska and western British Columbia (17). This diversity of provenance suggests that sediments from the northern Gulf of Alaska integrate ice-rafted detritus from the western Cordilleran ice streams, carried northward and westward by the Alaska Current and Alaska Coastal Current. Meltwater influx likely accelerates these current systems during episodes of rapid ice retreat (22). The youngest Siku event (S1) captured in the U1419 marine IRD record coincides with terrestrial cosmogenic-exposure dates for retreat of the western Cordilleran, roughly synchronous on the marine margin from Alaska to southern British Columbia (23, 24).

¹College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, OR, USA. ²Australian National University, Canberra ACT. ³Department of Geological and Environmental Sciences, Appalachian State University, Boone, NC, USA. ⁴United States Geological Survey, Corvallis, OR, USA. ⁵Department of Earth Sciences, Institute of Geochemistry and Petrology, ETH Zurich, Zurich, Switzerland. ⁶United States Geological Survey, Menlo Park, CA, USA.

⁷University of Central Missouri, Warrensburg, MO, USA.

*Corresponding author. Email: mo.walczak@oregonstate.edu

†These authors contributed equally to this work.

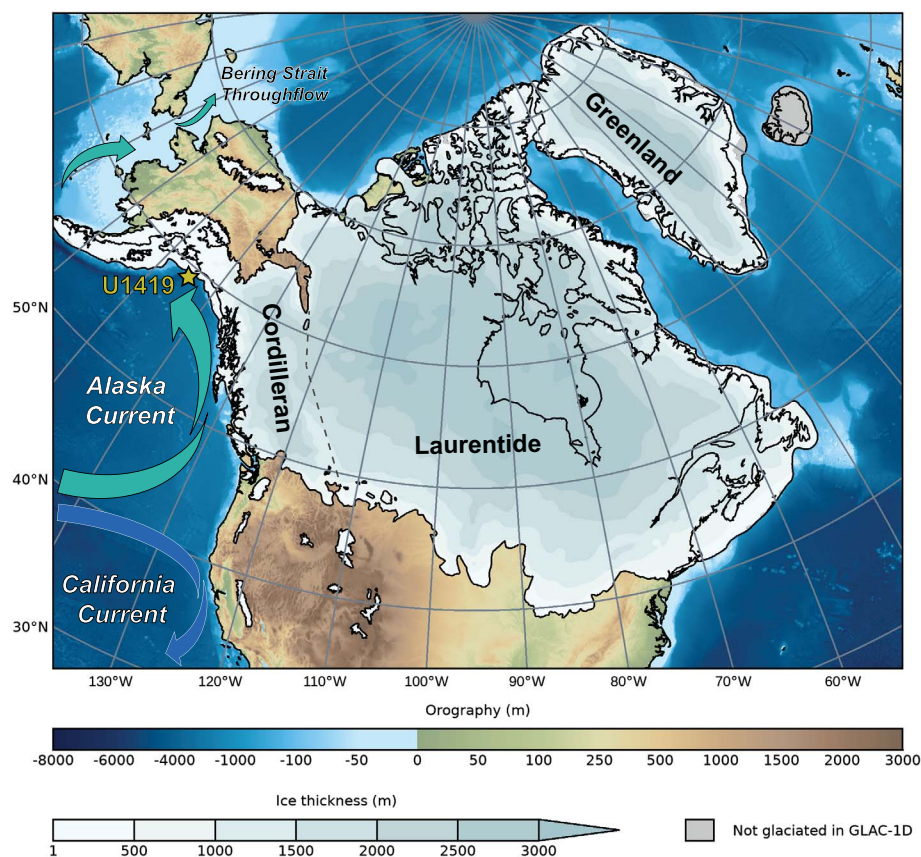
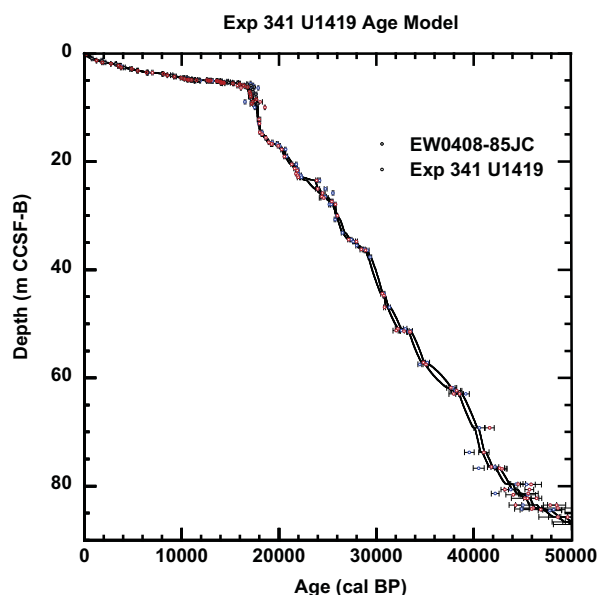


Fig. 1. Map showing study area. Location of International Ocean Discovery Program Expedition 341 Site U1419 (59°31.9'N, 144°8.0'W; 698-m depth) is shown on modern ETOPO1 bathymetry (58). The GLAC-1D modeled LGM thickness of the Cordilleran, Laurentide, and Greenland ice sheets is shown (59), with margins constrained by the proxy-derived LGM ice extent (60). The North American ice sheet saddle separating the Cordilleran from the Laurentide is delineated as a dashed black line. Approximate directions and extent of major northeast Pacific surface currents, including the Alaska Current, California Current, and Bering Strait Throughflow are also shown.

Fig. 2. Bayesian age model for Expedition 341 Site U1419 (see materials and methods) based on the Marine13 calibration curve. The solid lines are the $\pm 1\sigma$ error envelope on the age model, shown versus adjusted composite depth below seafloor (CCSF-B) (17). All calibrated planktic (blue) and benthic (red) radiocarbon dates are shown. Data from core EW0408-85JC are filled circles (28, 33), whereas data from U1419 are open circles. See materials and methods for Marine20 age model.



The modern water mass at the depth of Site U1419 is at the approximate boundary of North Pacific Intermediate Water [NPIW; an intermediate water mass partially ventilated to the atmosphere in the Sea of Okhotsk (25)] and Pacific Deep Water [PDW; a primary water mass sourced around Antarctica as Circumpolar Deep Water with relatively high preformed ^{14}C age, transited northward from the Southern Ocean near the seafloor as Antarctic Bottom Water and returned southward at mid-depths, mixing with overlying waters but without interacting with the atmosphere (26)]. A ^{14}C -depleted water mass expanded through much of the deep Pacific during the most recent glacial termination (27), with high apparent ages at intermediate depths approximately coeval with two events of increased benthic reservoir age observed in the Gulf of Alaska (28). These events, associated with deglacial increases in atmospheric CO_2 , have been inferred to reflect the upward mixing and redistribution of a ^{14}C -depleted deep water mass (29, 30). Evidence from authigenic ϵNd provides further support for the interpretation of an increased contribution of abyssal Pacific waters to intermediate-depth North Pacific waters at these times (31, 32).

We measured 82 benthic-planktic (B-P) radiocarbon pairs spanning the past ~50,000 yr B.P. at Site U1419, overlapping with and extending 28 paired measurements from <18,000 calibrated ^{14}C years before present (cal yr B.P.) in colocated core EW0408-85JC [59°33.32'N, 144°9.21'W, 682 m (28, 33)] (fig. S2). At this site, an increase in the radiocarbon age of bottom waters relative to the surface (e.g., high B-P values) may indicate (i) decreased circulation rate increasing the true age of subsurface waters; (ii) decreased gas exchange of source waters imparting an apparently high preformed age on the watermass; (iii) shoaling or increased mixing of underlying ^{14}C -deficient deep waters with overlying intermediate waters; and/or (iv) decreased mixing of intermediate waters with the sea surface [for example, through enhanced shallow stratification (34)]. The largest B-P age excursions observed here cannot be driven solely by decreases in surface age (for example, by less mixing with underlying older waters or greater effective local gas exchange), as this would require implausible surface-water ages younger than that of the coeval atmosphere (28). Thus, substantial increases of B-P age differences must reflect increases in subsurface reservoir ages.

We limit the evaluation of the B-P at Site U1419 to ages <42,000 cal yr B.P. (76.5 m CCSF-B), where analytical precision is sufficient for meaningful interpretation (see materials and methods). The B-P age differences over that interval average 790 ± 340 (1σ) years,

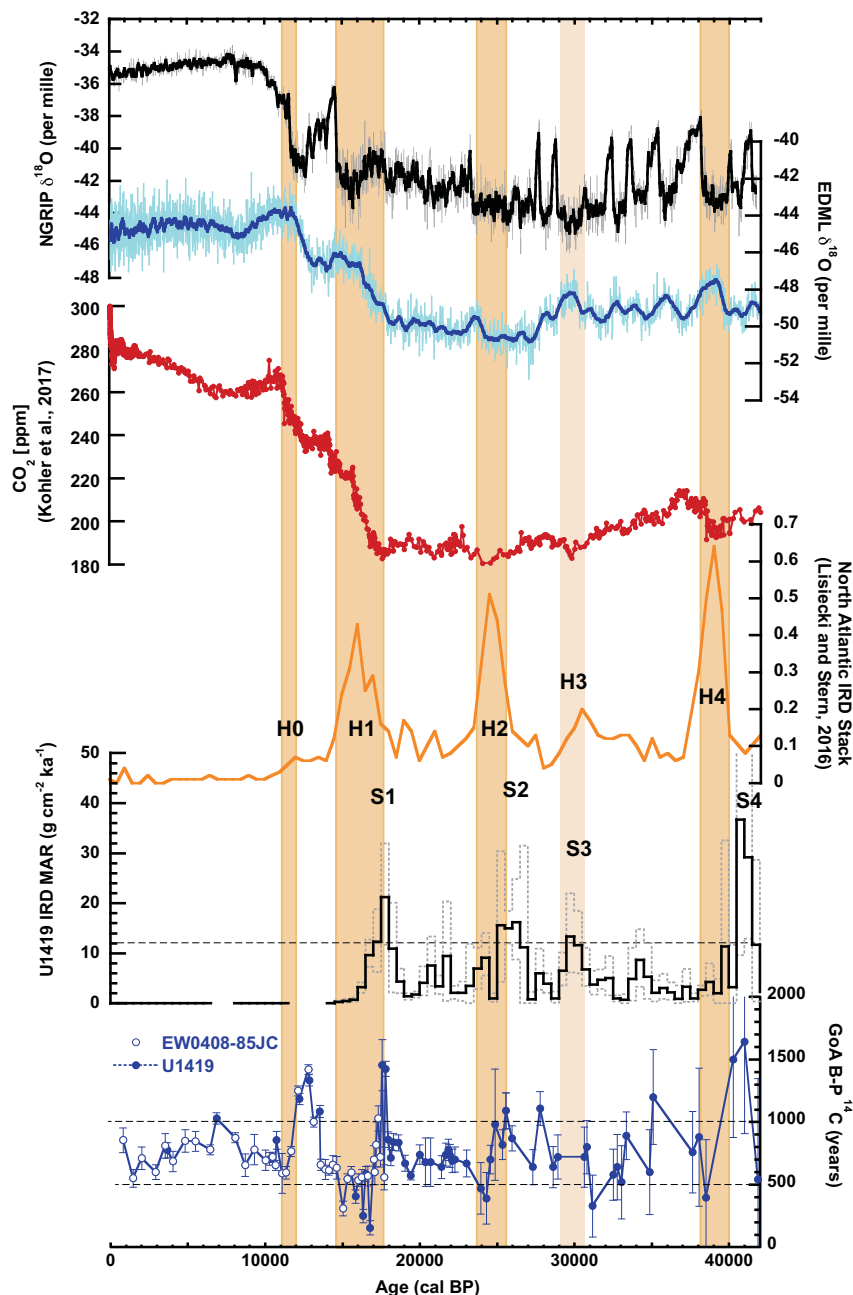


Fig. 3. Global records of climate changes during the latest Pleistocene. Greenland [$\delta^{18}\text{O}$ from North Greenland Ice Core Project (NGRIP), black; 100-year smoothing shown in bold] (61) and Antarctica ($\delta^{18}\text{O}$ from EDML, blue; 100 year smoothing shown in bold) (62) are shown on the synchronized AICC2012 time scale (63). Global atmospheric CO_2 from ice cores plotted in red (64). Atlantic ice-rafted debris stack [normalized units in orange (65)] and U1419 IRD MAR calculated over 500-year increments (black) with $\pm 1\sigma$ uncertainty envelope (dashed gray lines). U1419 B-P ^{14}C (blue, $\pm 1\sigma$ uncertainty; EW0408-85JC data denoted by open symbols). Timing of the North Atlantic Heinrich events shown in light orange bars; H1 to H4 from synthesis of (65), H0 from (66). The dashed line on the U1419 IRD panel denotes the level of $12 \text{ g cm}^{-2} \text{ ka}^{-1}$ that define Siku events 1 to 4. The dashed straight lines on the U1419 B-P panel denote 1000 and 500 years; Siku events are associated with regional B-P ^{14}C age differences >1000 years, whereas Heinrich events are associated with values <500 years.

similar to (but with broader range than) the site's average Holocene ($<11,700$) value of 750 ± 120 (1σ) years (28), and estimated modern “pre-bomb” ^{14}C age differences between

surface and intermediate waters of 675 ± 90 (1σ) years (35). Over the past 42,000 cal yr B.P., four identifiable episodes occurred in which B-P age differences exceed 1000 years (con-

firmed by two or more B-P pairs), centered at 12,800, 17,700, 25,600, and 40,600 cal yr B.P. (Fig. 3). The older three of these episodes of high B-P are synchronous with Siku events 1, 2, and 4.

A smaller Siku event (S3), potentially associated with H3 near 30,000 cal yr B.P., has no accompanying B-P anomaly at Site U1419 (Fig. 3). However, in this interval, the resolution of the ^{14}C data set is low, so an event may have been missed. The youngest B-P excursion ($\sim 12,800$ cal yr B.P.), previously identified in nearby core EW0408-85JC (28), does not correspond to a Siku event as expressed in IRD, but it is associated with terrestrial ice retreat (36–38), high meltwater fluxes from land, and adjacent surface ocean cooling (33, 39). By this time, the Cordilleran outlet glaciers had retreated into silled fjords (36) that served as sediment traps, leaving little expression of enhanced sediment or IRD fluxes in the open ocean (40).

The North Pacific may be sensitive to the Asian monsoon, which provides a net freshwater source that contributes to upper ocean stratification, limiting subsurface ventilation (41). High B-P and Siku events are initiated during strong monsoon intervals (higher net freshwater flux to the North Pacific), whereas low B-P events tend to occur in concert with weak monsoons (which are coincident with Atlantic Heinrich events; Fig. 4). Northward heat transport associated with strong monsoons may contribute to net negative mass balance and retreat of the marine margin of the Cordilleran Ice Sheet (42). In turn, freshwater discharge from the Cordilleran Ice Sheet can further stratify the northeast Pacific, allow intermediate water reservoir ages to rise as vertical mixing with the surface ocean is suppressed and mixing with the deep ocean is favored, while triggering dynamic responses in the North Atlantic sector (14, 39, 43, 44). Chinese Cave $\delta^{18}\text{O}$ (9) leads Siku events by ~ 1000 to 3000 years (Fig. 4), reasonable for an ice sheet response to warming. This would seemingly support the hypothesis that low-latitude processes are an important driver of high-latitude climate (45).

Antarctic warming, as reconstructed from ice core $\delta^{18}\text{O}$, follows within ~ 1000 years of Siku events, with peak warmth ~ 2500 years after peak IRD MAR (Figs. 3 and 5). This phasing appears to preclude northward propagation of an Antarctic trigger for Siku events (3) but could support a mechanism in which Cordilleran meltwater cools the North Pacific and in turn the North Atlantic, whereby reduction in AMOC triggers Antarctic warming (43). Changes in Southern and Pacific Ocean circulation associated with Antarctic warming during the Heinrich events are hypothesized to be related to the release of carbon from the interior ocean, driving some portion of

Fig. 4. Paleoclimate connection of the equatorial and high-latitude North Pacific. The B-P ^{14}C record of U1419 (solid blue dots) and its site survey core EW0408-85JC (open blue dots), reflecting changes in the circulation and/or ventilation structure of the intermediate-upper Pacific Ocean, superimposed on the U-Th dated Chinese speleothem $\delta^{18}\text{O}$ record interpreted as reflecting strength of the Asian monsoon (green dashed line) (9). Episodes of high B-P at U1419, and attendant instability of the Cordilleran (Fig. 4), appear to track periods of strong Asian monsoon. Heinrich events, anomalously low B-P ^{14}C differences, and periods of weak Asian monsoon follow. Timing of the North Atlantic Heinrich events shown in light orange bars; H1 to H4 from synthesis of (65), H0 from (66).

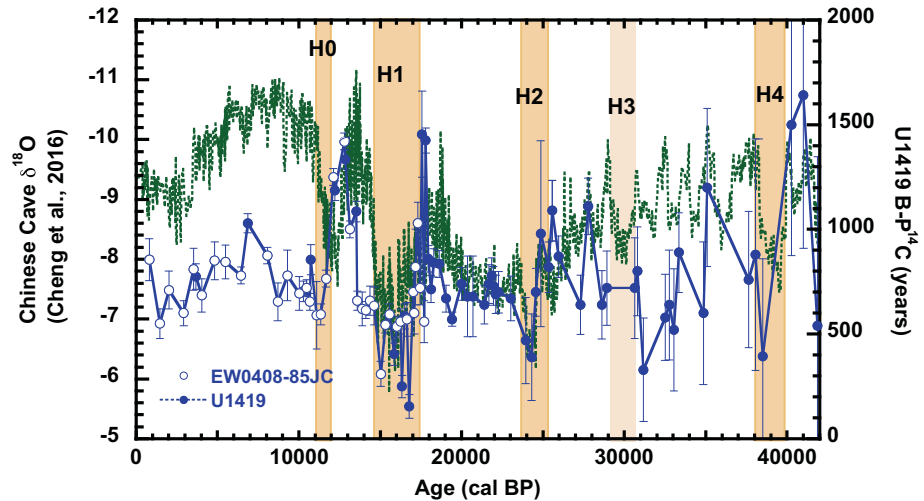
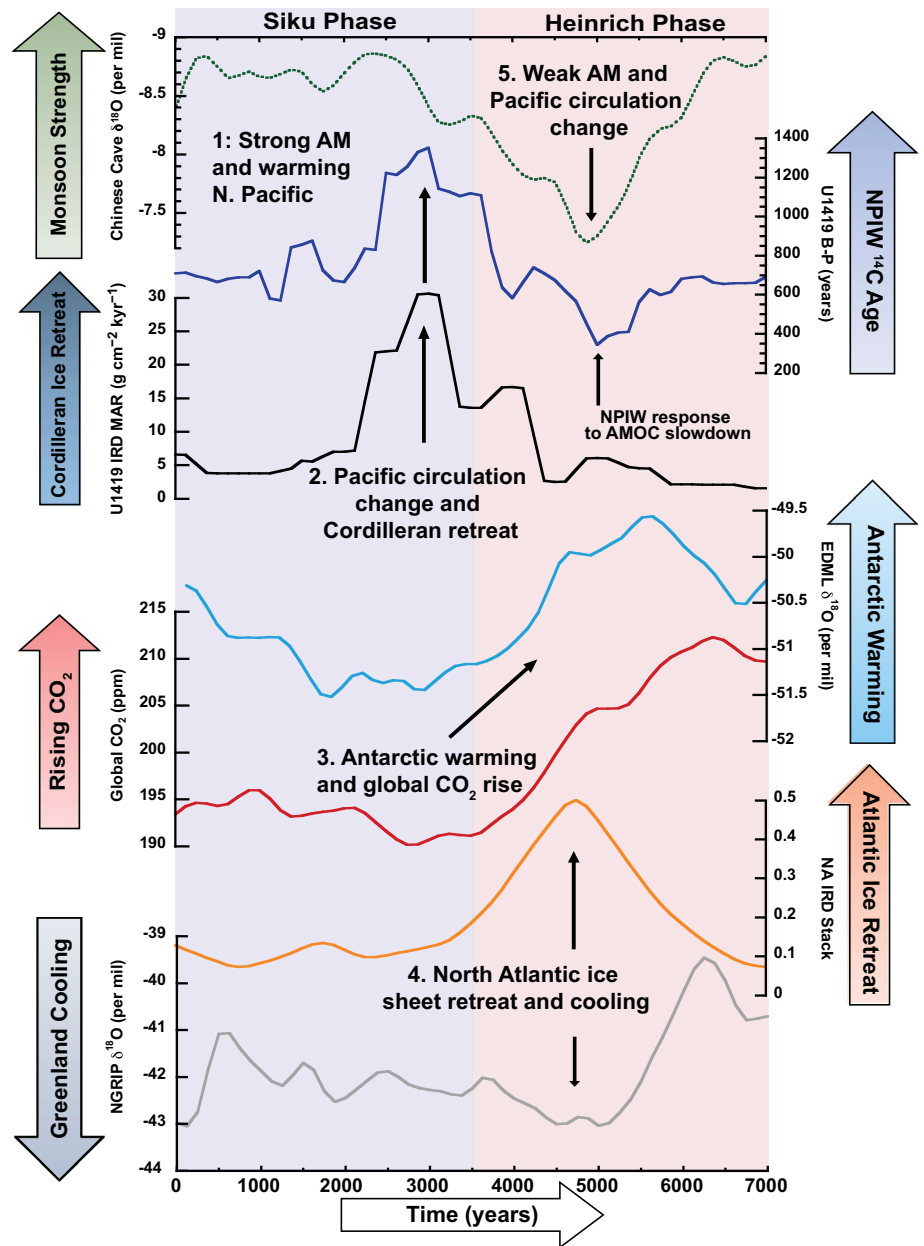


Fig. 5. The progression of global climate anomalies surrounding retreat of the Northern Hemisphere ice sheets over the past ~45,000 cal yr B.P. Canonical events here are averages of the intervals around well-resolved Siku events 1, 2, and 4 aligned on their published chronologies according to the timing of maxima in ice-rafted debris MAR in Site U1419. Each record was then interpolated to 125-year time step and smoothed with a 750-year Gaussian filter. The 7000 years of canonical changes surrounding each event are shown for the Asian monsoon [AM; Chinese Cave speleothem $\delta^{18}\text{O}$ (9)], Cordilleran ice-rafted debris (IRD) and Northeast Pacific Intermediate Water (NPIW) ventilation age as reconstructed from Site U1419, Antarctic temperature [$\delta^{18}\text{O}$ from EDML, blue (62)], global atmospheric CO_2 (64), North Atlantic IRD (65), and Greenland temperature [$\delta^{18}\text{O}$ from NGRIP (61)]. Ice core records are shown on the synchronized AICC2012 time scale (63). Early expression of Northern Hemisphere ice sheet instability (i.e., the “Siku Phase”) is seen in the Pacific sector, where intensified Asian monsoon (9, 45) coincides with the start of northeast Pacific ice rafting, and to large positive excursions in the ^{14}C age of NPIW, followed by Antarctic warming and global CO_2 rise. In the “Heinrich Phase,” the marine-terminating ice margins of the North Atlantic retreat, generating Heinrich events, and Greenland cools. The Asian monsoon weakens and NPIW ^{14}C ages decrease, likely as a response to a reduction in AMOC (3, 9, 55).



the following increase in atmospheric CO₂ (32, 46) (Fig. 5).

Atlantic Heinrich events lag Pacific Siku events by 1370 ± 550 years for peak IRD. This phasing precludes teleconnection mechanisms in which Atlantic processes trigger the earlier expression of Pacific millennial-scale climate instability (Fig. 5 and fig. S10). The correspondence of high North Pacific B-P age differences with Siku events similarly precludes their origin as a downstream reduction of Pacific intermediate-water ventilation in response to AMOC suppression during Heinrich events.

Although high B-P age differences in the North Pacific precede Heinrich events, the lowest B-P age differences at Site U1419 co-occur with H0, H1, H2, and H4 and may plausibly reflect the regional response to Heinrich events (Fig. 3). Similar timing for low B-P is found in the northwest Pacific and Bering Sea (47–51), and may support a hypothesized interocean “see-saw” effect in shallow subsurface ventilation (52–54). Earth system model results suggest that freshwater input in the North Atlantic during Heinrich events could drive invigorated NPIW production and/or enhanced stratification between NPIW and PDW, both of which would contribute to younger intermediate water age (55, 56).

We have documented correspondence of increased northeast Pacific intermediate water ¹⁴C ages and ice-rafted debris delivery during the past 42,000 years. A finely resolved chronology implicates Cordilleran ice retreat (Siku events) as early in a chain of prominent climate events, following strong Asian monsoons but leading the Heinrich events of the North Atlantic, Antarctic warming, and global CO₂ rise (Fig. 5). These observations indicate that the Cordilleran Ice Sheet and North Pacific Ocean play an active role in millennial-scale climate oscillations and preclude direct forcing of Cordilleran ice retreat from Heinrich events. Mechanisms linking Pacific and Atlantic ice retreat events may include atmospheric heat transports and adjustments in the Arctic (44), net freshwater transports from the Pacific to the Atlantic when Bering Strait is open (39), and/or rapid sea-level rises accompanying ice melt. Precursor retreat of the smaller European ice sheets, posited as a trigger for Laurentide ice purges (8), has since been discounted (1). However, the early and rapid Cordilleran ice losses documented here are more likely to influence Atlantic ice sheet retreat. The relative response times of these various interacting systems range from subdecadal (monsoons and atmospheric adjustments), to centuries (smaller Cordilleran ice and water-mass ventilation), to millennia (larger Laurentide ice and interbasin deep circulation), offering a possibility that linkages across a range of time scales could drive auto-oscillating behav-

ior when ice is present as a triggering mechanism (57).

REFERENCES AND NOTES

- S. R. Hemming, *Rev. Geophys.* **42**, RG1005 (2004).
- W. S. Broecker, *Nature* **372**, 421–424 (1994).
- P. U. Clark, S. W. Hostetler, K. J. Meissner, N. G. Pisias, A. Schmittner, in *Ocean Circulation: Mechanisms and Impacts - Past and Future Changes of Meridional Overturning*, A. Schmittner, J. C. H. Chiang, S. R. Hemming, Eds., Geophysical Monograph 173 (American Geophysical Union, 2007), pp. 209–246.
- G. Shaffer, S. M. Olsen, C. J. Bjerrum, *Geophys. Res. Lett.* **31**, L24202 (2004).
- S. A. Marcott et al., *Proc. Natl. Acad. Sci. U.S.A.* **108**, 13415–13419 (2011).
- J. N. Bassis, S. V. Petersen, L. Mac Cathles, *Nature* **542**, 332–334 (2017).
- D. R. MacAyeal, *Paleoceanography* **8**, 775–784 (1993).
- G. C. Bond, R. Lotti, *Science* **267**, 1005–1010 (1995).
- H. Cheng et al., *Nature* **534**, 640–646 (2016).
- T. Blunier, E. J. Brook, *Science* **291**, 109–112 (2001).
- A. T. Hewitt, D. McDonald, B. D. Bornhold, *Geophys. Res. Lett.* **24**, 3261–3264 (1997).
- A. C. Mix, D. C. Lund, N. G. Pisias, P. Boden, L. Bornmalm, M. Lyle, J. Pike, in *Mechanisms of Global Climate Change at Millennial Time Scales*, Geophysical Monograph 112 (American Geophysical Union, 1999), pp. 127–148.
- I. L. Hendy, T. Cosma, *Paleoceanography* **23**, PA2101 (2008).
- E. Maier et al., *Nature* **559**, 241–245 (2018).
- J. M. Swartz, S. P. S. Gulick, J. A. Goff, *Geochem. Geophys. Geosyst.* **16**, 165–177 (2015).
- J. Seguinot, I. Rogozhina, A. P. Stroeven, M. Margold, J. Kleman, *Cryosphere* **10**, 639–664 (2016).
- J. M. Jaeger et al., and the Expedition 341 Scientists (Integrated Ocean Drilling Program, College Station, TX, 2014). doi:10.2204/iodp.pr.341.2014.
- P. J. Reimer et al., *Radiocarbon* **55**, 1869–1887 (2013).
- T. J. Heaton et al., *Radiocarbon* **62**, 779–820 (2020).
- S. J. Marshall, M. R. Koutnik, *Paleoceanography* **21**, PA2021 (2006).
- M. N. Koppes, B. Hallet, *Geology* **30**, 47–50 (2002).
- T. C. Royer, B. Finney, *Oceanography (Wash. D.C.)* **33**, (2020).
- C. M. Darvill, B. Menounos, B. M. Goehring, O. B. Lian, M. W. Caffee, *Geophys. Res. Lett.* **45**, 9710–9720 (2018).
- A. J. Lesnek, J. P. Briner, C. Lindqvist, J. F. Baichtal, T. H. Heaton, *Sci. Adv.* **4**, eaar5040 (2018).
- L. D. Talley, *J. Phys. Oceanogr.* **23**, 517–537 (1993).
- L. D. Talley, *Oceanography (Wash. D.C.)* **26**, 80–97 (2013).
- T. M. Marchitto, S. J. Lehman, J. D. Ortiz, J. Flückiger, A. van Geen, *Science* **316**, 1456–1459 (2007).
- M. Davies-Walczak et al., *Earth Planet. Sci. Lett.* **397**, 57–66 (2014).
- E. L. Sikes, C. R. Samson, T. P. Guilderson, W. R. Howard, *Nature* **405**, 555–559 (2000).
- L. Skinner et al., *Earth Planet. Sci. Lett.* **411**, 45–52 (2015).
- C. Basak, E. E. Martin, K. Horikawa, T. M. Marchitto, *Nat. Geosci.* **3**, 770–773 (2010).
- J. Du, B. A. Haley, A. C. Mix, M. H. Walczak, S. K. Praetorius, *Nat. Geosci.* **11**, 749–755 (2018).
- M. H. Davies et al., *Paleoceanography* **26**, PA2223 (2011).
- S. Khaliwala, F. Primeau, M. Holzer, *Earth Planet. Sci. Lett.* **325–326**, 116–125 (2012).
- R. M. Key et al., *Global Biogeochem. Cycles* **18**, GB4031 (2004).
- B. Menounos et al., *Science* **358**, 781–784 (2017).
- J. P. Briner, D. S. Kaufman, *J. Quat. Sci.* **23**, 659–670 (2008).
- D. S. Kaufman, R. Scott Anderson, F. S. Hu, E. Berg, A. Werner, *Quat. Sci. Rev.* **29**, 1445–1452 (2010).
- S. K. Praetorius et al., *Sci. Adv.* **6**, eaay2915 (2020).
- E. A. Cowan et al., *Bull. Geol. Soc. Am.* **122**, 1067–1080 (2010).
- J. Emile-Geay, *J. Geophys. Res.* **108**, 3178 (2003).
- S. Hostetler, N. Pisias, A. Mix, *Quat. Sci. Rev.* **25**, 1168–1185 (2006).
- D. M. Roche, A. P. Wiersma, H. Renssen, *Clim. Dyn.* **34**, 997–1013 (2010).
- S. Praetorius, M. Rugenstein, G. Persad, K. Caldeira, *Nat. Commun.* **9**, 3124 (2018).
- J. W. Beck et al., *Science* **360**, 877–881 (2018).
- M. Sarnthein, B. Schneider, P. M. Grootes, *Clim. Past* **9**, 2595–2614 (2013).
- N. Ahagon, *Geophys. Res. Lett.* **30**, 2097 (2003).
- Y. Okazaki et al., *Science* **329**, 200–204 (2010).
- L. Max et al., *Clim. Past* **10**, 591–605 (2014).
- J. W. B. Rae et al., *Paleoceanography* **29**, 1–23 (2014).
- W. R. Gray et al., *Nat. Geosci.* **11**, 1 (2018).
- D. C. Lund, *Earth Planet. Sci. Lett.* **381**, 52–62 (2013).
- A. T. Kotilainen, N. J. Shackleton, *Nature* **377**, 323–326 (1995).
- O. A. Saenko, A. Schmittner, A. J. Weaver, *J. Clim.* **17**, 2033–2038 (2004).
- M. O. Chikamoto et al., *Deep. Res. Part II Top. Stud. Oceanogr.* **61–64**, 114–126 (2012).
- X. Gong et al., *Nat. Commun.* **10**, 656 (2019).
- M. Crucifix, *Philos. Trans. A Math. Phys. Eng. Sci.* **370**, 1140–1165 (2012).
- C. Amante, B. W. Eakins, ETOPO1 Global Relief Model converted to PanMap layer format. NOAA-National Geophysical Data Center, PANGAEA (2009). <https://doi.org/10.1594/PANGAEA.769615>.
- W. R. Peltier, D. F. Argus, R. Drummond, *J. Geophys. Res. Solid Earth* **120**, 450–487 (2015).
- R. McNeely, A. S. Dyke, J. R. Southon, *Geol. Surv. Canada Open File* **5049** (2006).
- K. K. Andersen et al., *Nature* **431**, 147–151 (2004).
- L. Augustin et al., *Nature* **429**, 623–628 (2004).
- D. Veres et al., *Clim. Past* **9**, 1733–1748 (2013).
- P. Köhler, C. Nehrbass-Ahles, J. Schmitt, T. F. Stocker, H. Fischer, *Earth Syst. Sci. Data* **9**, 363–387 (2017).
- L. E. Lisiecki, J. V. Stern, *Paleoceanography* **31**, 1368–1394 (2016).
- C. Pearce et al., *Paleoceanography* **30**, 1613–1624 (2015).

ACKNOWLEDGMENTS

We are grateful to the crew and shipboard science party of IODP Expedition 341, as well as the curators at the IODP Gulf Coast Repository. **Funding:** This work was supported by the Australian Research Council (FS100100076), the Australia-New Zealand IODP Commission, the U.S. Science Support Program and Consortium for Ocean Leadership, the National Science Foundation (0242084, 0602395, 0728315, 1103538, 1204204, 1357529, 1360894, 1434945, 1436903, 1502754, 1924215 1929486), and the American Australian Association. **Author contributions:** M.W., A.M., and S.F. designed study, generated ¹⁴C data, and analyzed datasets. E.C. generated and aided in interpretation of IRD data. K.F. generated dates for oldest ¹⁴C samples. J.P., T.H., and J.D. prepared samples for ¹⁴C analysis. S.Z. prepared species-specific benthic ¹⁴C samples for evaluation of down-slope transport. J.A. contributed to interpretation of modeled ice sheet behavior and figure creation, including Cordilleran meltwater input. A.S. contributed to modeling of ocean radiocarbon reservoir age. M.W., A.M., S.F., E.C., S.P., J.S., J.D., B.H., and S.Z. contributed to writing of the paper. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6517/716/suppl/DC1
Materials and Methods
Figs. S1 to S15
Tables S1 to S10
References (67–77)

26 December 2019; accepted 17 September 2020
Published online 1 October 2020
10.1126/science.aba7096

ECOLOGICAL GENOMICS

Genomic architecture of a genetically assimilated seasonal color pattern

Karin R. L. van der Burg^{1*}, James J. Lewis^{1,2}, Benjamin J. Brack¹, Richard A. Fandino¹, Anyi Mazo-Vargas¹, Robert D. Reed^{1*}

Developmental plasticity allows genomes to encode multiple distinct phenotypes that can be differentially manifested in response to environmental cues. Alternative plastic phenotypes can be selected through a process called genetic assimilation, although the mechanisms are still poorly understood. We assimilated a seasonal wing color phenotype in a naturally plastic population of butterflies (*Junonia coenia*) and characterized three responsible genes. Endocrine assays and chromatin accessibility and conformation analyses showed that the transition of wing coloration from an environmentally determined trait to a predominantly genetic trait occurred through selection for regulatory alleles of downstream wing-patterning genes. This mode of genetic evolution is likely favored by selection because it allows tissue- and trait-specific tuning of reaction norms without affecting core cue detection or transduction mechanisms.

Developmental plasticity—the ability of one genotype to produce alternative phenotypes in response to environmental cues—is an important driver of phenotypic diversity (1, 2). Because plasticity allows genomes to encode different potential phenotypes, it can be a source of phenotypic variation that natural selection can act upon (3). Waddington coined the term “genetic assimilation” to describe a reduction of plasticity, where selection leads to expression of an ancestrally conditional phenotype in the absence of the original environmental cue (4). Little is currently known, however, about the genetic mechanisms that underlie the assimilation of ancestrally plastic phenotypes (5).

We sought to characterize the genomic architecture of genetic assimilation using seasonal wing color plasticity in the common buckeye butterfly *Junonia coenia*, which develops a pale tan wing color under warm, long-day conditions, and a dark red color under cold, short-day conditions (Fig. 1A) (6). The wing color reaction norm can evolve through both artificial and natural selection, and both tan and red phenotypes are fixed in some populations and closely related species (7, 8).

To isolate alleles that control wing color plasticity, we generated two artificial selection lines for increased and reduced wing color plasticity, respectively. Butterflies selected for increased plasticity (the “Plastic” line) were subjected to Waddington’s alternating selection regime (9): When reared under warm, long-day conditions, the palest butterflies were

selected to breed, after which offspring were reared under cool, short-day conditions, and the reddest butterflies were selected. Within six generations, the color difference between warm- and cold-reared animals increased (Fig. 1B and fig. S1), and the reaction norm slope increased accordingly (Fig. 1, C and D, and table S1). Selection for reduced wing color plasticity (the “Red” line) was achieved by rearing offspring under warm conditions and selecting the reddest butterflies each generation. After 12 generations, the red phenotype became predominant under conditions where butterflies would normally be tan (Fig. 1E and fig. S1), and there was a significant reduction in the reaction norm slope (Fig. 1, C and D, and table S1). Thus, we created a Plastic line in which the red wing phenotype is induced by environmental cues, and a Red line in which the red wing phenotype is assimilated.

Evolution of environmental response should be attributable to changes in one of three mechanisms: (i) cue detection, the capacity to detect differences in the environment; (ii) cue transduction, the translation of a cue into an internal signal such as a hormone titer; and (iii) tissue response, the way a specific tissue responds to the internal signal (10). We thus sought to determine to what extent we could explain the plasticity differences between our selection lines through changes in any of these mechanisms.

Wing color plasticity in *J. coenia* is caused by environmentally induced changes in the timing of an ecdysone pulse during early pupal development (11). If assimilation of the red phenotype occurred because the ability to detect or transduce cues was lost, we would predict that warm-reared Red line butterflies would have an ecdysone profile similar to that

of cold-reared Plastic line butterflies (which are red). However, we found that ecdysone titers in warm-reared butterflies from the Red line were the same as in warm-reared butterflies from the Plastic line (Fig. 1F), which are red and tan, respectively. Ecdysone titers responded to environmental cues similarly in both lines. This indicates that in the genetically assimilated Red line, wings became red despite a normally functioning cue detection and transduction system. Thus, we infer that genetic assimilation of red wing coloration occurred through changes in how wing tissues respond to an otherwise conserved endocrine signal.

To determine the genetic basis of changes in wing color plasticity, we performed F_3 crosses using individuals from the Red and Plastic lines and scored phenotypes of individuals raised under conditions that would normally induce tan phenotypes (i.e., warm, long light cycle). We then selected 20 of the most extreme tan (i.e., environment-responsive) and 21 of the most extreme red (i.e., environment-unresponsive) individuals for whole-genome resequencing (tables S2 and S3). A genome-wide association analysis, using the amount of red color in warm-reared animals to signify reduced plasticity, revealed four genomic loci (Fig. 2A). We identified candidate genes at these loci by comparing hindwing mRNA levels between the Red and Plastic lines at four developmental stages spanning before, during, and after the pupal ecdysone signal that determines wing color (fig. S5A). We detected 13 differentially expressed genes (Wald adjusted $P < 0.05$) within 25 kb of a plasticity-associated nucleotide variant (Fig. 2, B and C), of which the majority were differentially expressed late in development, during pigment synthesis (Fig. 2C).

We next used one of three additional criteria to select a subset of these genes for functional validation: (i) suspected wing-patterning function (*cortex*), (ii) highest expression levels at the differentially expressed stage (*trehalase* and *Dscam3*), or (iii) differential expression across multiple developmental stages (*CG8930*, hereafter *herfst*). We then used CRISPR/Cas9 somatic deletion mosaics (figs. S8 and S9) to test the function of *Dscam3* in the Plastic line, and *cortex*, *trehalase*, and *herfst* in the Red line, consistent with which lines showed elevated expression of the respective targets. We did not observe *Dscam3* phenotypes, but all three genes in the Red line produced distinct mutant clones of tan scales in a background of otherwise normal Red line red scales. These are phenotypes where individual butterflies display wing patterns that are mosaic for both summer and autumn coloration (Fig. 2D). One of the genes, *cortex*, is implicated in wing pattern adaptation in peppered moths and *Heliconius* butterflies (12, 13). The others are

¹Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY, USA. ²Baker Institute for Animal Health, Cornell University, Ithaca, NY, USA.

*Corresponding author. Email: krv32@cornell.edu (K.R.L.v.d.B.); robertreed@cornell.edu (R.D.R.)

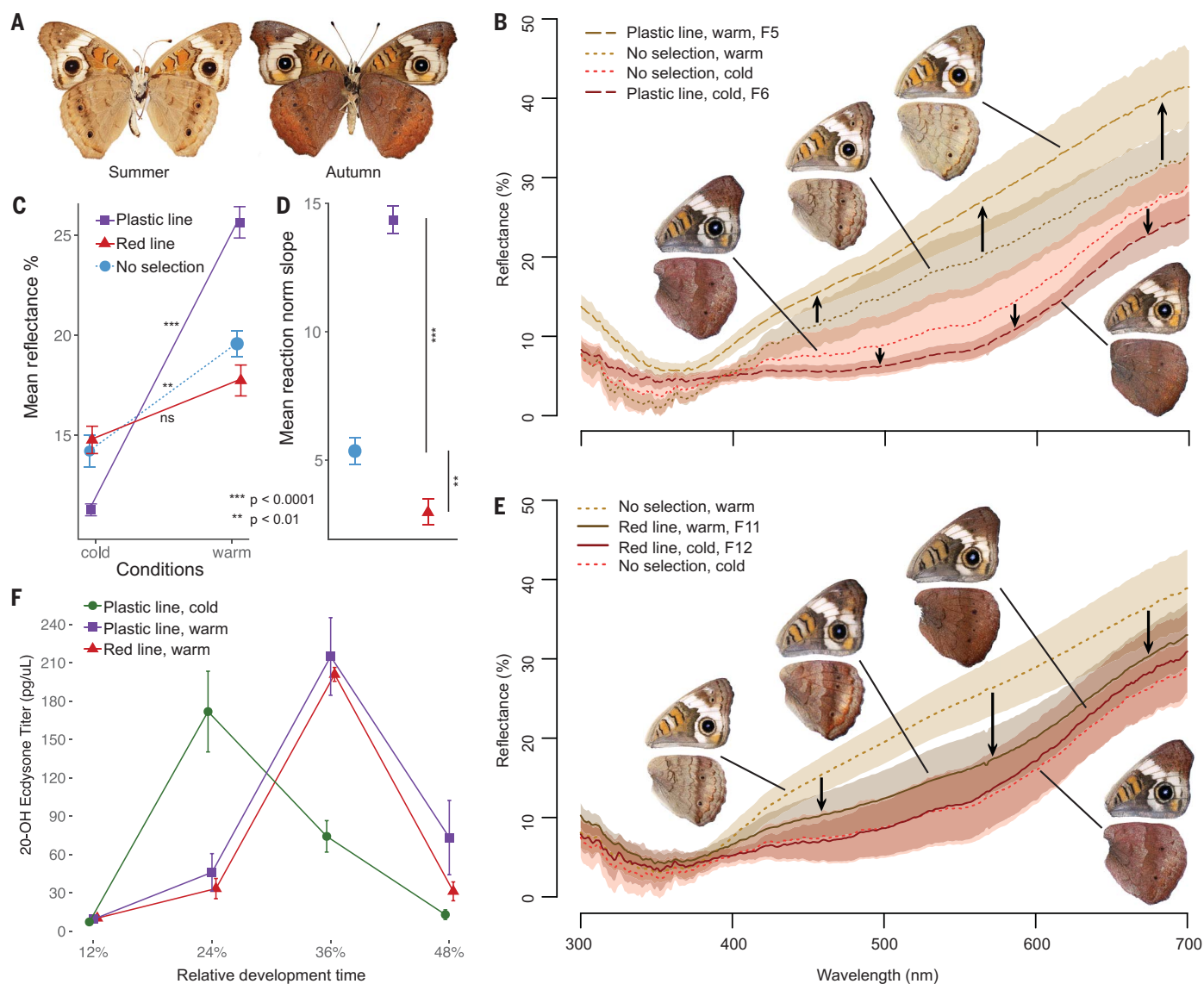


Fig. 1. Rapid selection for increased and reduced plasticity. (A) Seasonal morphs of *J. coenia*. (B) Wing color response differences after six generations of selection for increased plasticity (warm = 27°C and 16 hours of light, cold = 19°C and 8 hours of light). (C and D) Wing color reaction norms (C) and slopes (D) differ between selection lines (t test P values shown). (E) Wing color response differences after 12 generations of selection for reduced plasticity. (F) 20-hydroxy-ecdysone titers for the different selection lines under different conditions show no difference between warm-reared Plastic line and Red line butterflies. Error bars in (C), (D), and (F) and shaded areas in (B) and (E) denote SD.

trehalase, which encodes an enzyme that converts trehalose into glucose, a constituent of the red wing pigment rhodommatin (14), and *herfst*, a previously uncharacterized zinc finger domain gene. In sum, knockouts of three candidate genes showed that each is individually necessary for production of the red autumn phenotype, and the loss of function of any one of them phenocopies the tan summer phenotype.

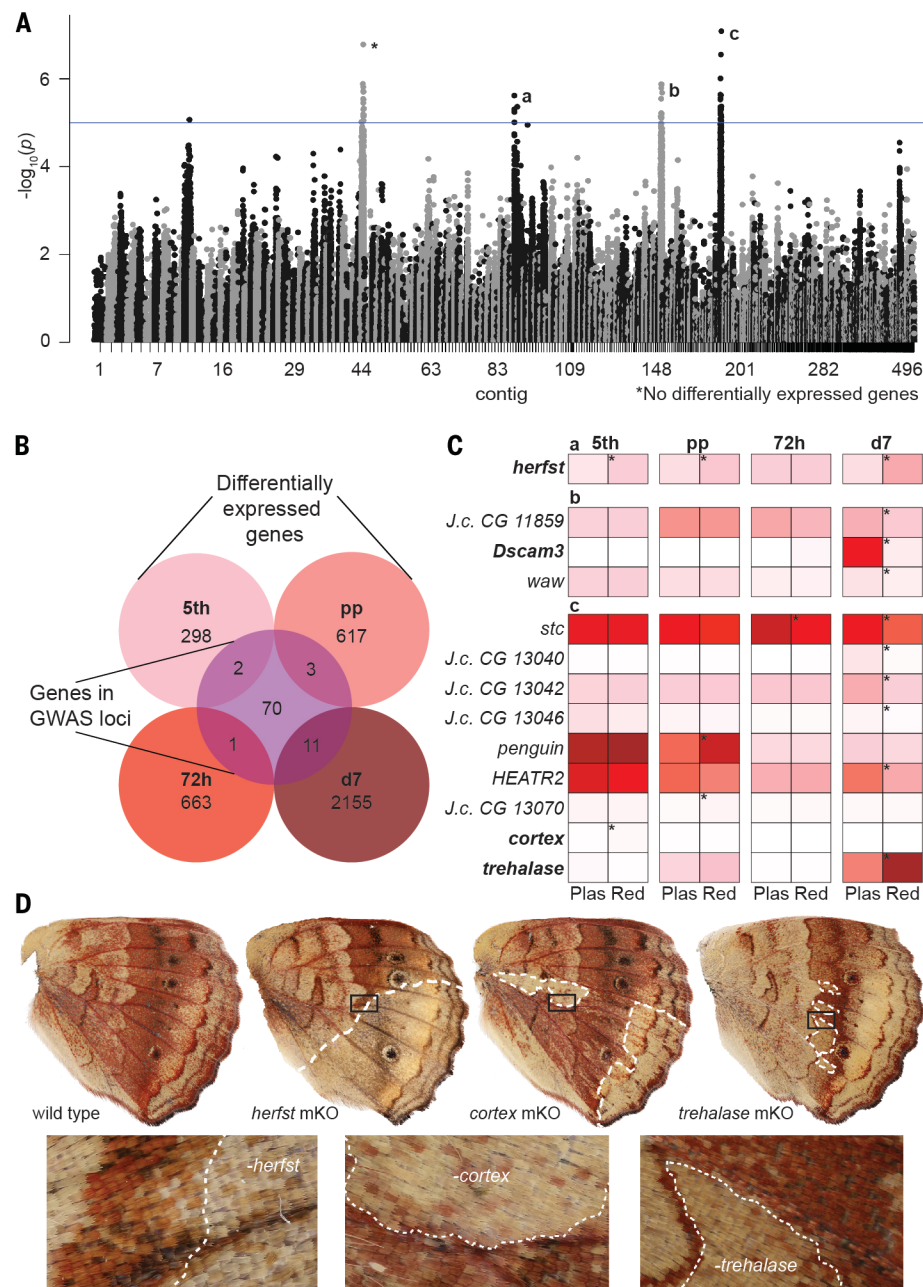
The evolution of reaction norms might be traceable to differences in regulatory regions that control the responsiveness of a gene to an inductive signal, such as a hormone (15). This model is consistent with mRNA expression differences observed between our selection lines, coupled with the lack of protein coding variation. To further characterize the

nature of cis-regulatory variation between selection lines, we used ATAC-seq (assay for transposase-accessible chromatin with high-throughput sequencing) to identify active cis-regulatory elements (CREs), and Hi-C chromatin contact mapping to assess physical interactions between CREs and gene promoters (16). Genome-wide chromatin accessibility was similar between lines, with only 13% of ATAC-seq peaks differing (Wald adjusted $P < 0.05$), most of which (11%) varied late in development during pigment deposition. Of 64 associated nucleotide variants, 58 were located in presumptive CREs (i.e., within ATAC-seq peaks), whereas only two were nonsynonymous amino acid substitutions.

For *cortex*, ATAC-seq plus Hi-C highlighted a cluster of significant nucleotide variants in

the first intron that correspond to a series of nonvariable ATAC-seq peaks that physically interact with the *cortex* promoter (Fig. 3A, fig. S7, and table S9). For *trehalase*, ATAC-seq plus Hi-C revealed a single major ~40-kb upstream CRE varying in accessibility between lines (Wald adjusted $P < 0.05$) with strong promoter interaction during pigment synthesis (Fig. 3B, fig. S7, and table S9), suggesting differences in CRE activity. For *herfst*, ATAC-seq plus Hi-C revealed a hub of interactions where three CREs, including the *herfst* promoter, loop to each other. Two of these elements also showed differences in chromatin accessibility between lines (Wald adjusted $P < 0.05$) (Fig. 3C, fig. S7, and table S9), again suggesting line-specific differences in CRE activity. These results, coupled with

Fig. 2. Three genes underlying reaction norm changes between selection lines.
(A) Four loci showed multiple nucleotides significantly associated ($P \leq 0.00005$) with red color assimilation in Red line \times Plastic line F_3 crosses. Asterisk denotes an associated contig with no differentially expressed genes. (B) Differentially expressed genes across multiple stages [fifth instar (5th), pre-pupa (pp), 72 hours after pupation (72h), day 7 after pupation (d7)]. Overlap with center circle indicates genes within 50 kb of an associated nucleotide. GWAS, genome-wide association study. (C) Differential expression of candidate genes in Plastic and Red lines. Asterisks denote significance (Wald adjusted $P < 0.05$). Genes selected for functional validation are in bold. Letters a, b, and c correspond to nucleotide association regions shown in (A). (D) CRISPR/Cas9 somatic mosaic knockouts (mKO) confirm that *herfst*, *cortex*, and *trehalase* are all required for development of red pigmentation characteristic of the autumn morph.



observed expression differences, sequence associations in CREs, and lack of coding variation, support a model where genetic assimilation of red wing coloration was facilitated by selection on alleles of *cortex*, *trehalase*, and *herfst* with different cis-regulatory functions.

The three genes we identified in this study suggest mechanisms through which assimilation can evolve (Fig. 4). We speculate that *trehalase* plays a direct role in pigment synthesis, and that an environmentally insensitive increase of *trehalase* expression in the Red line promotes production of red pigmentation. It is noteworthy that *trehalase* is also up-regulated in cold-reared *J. coenia* (17), which suggests that this gene is involved in generating both

the induced and the assimilated red phenotype. In contrast to *trehalase*, differences in *cortex* and *herfst* expression occur during larval development, before the inductive ecdysone pulse (Fig. 4B). These pre-cue expression differences suggest that these genes play an upstream role in how the wing color regulatory network responds to the ecdysone signal. We speculate that early expression of these genes may tune the parameters of ecdysone response—for example, through threshold sensitivity and/or critical period. As with *trehalase*, higher mRNA abundance in the Red line suggests that this occurs through a loss of repression under summer-like conditions. Ultimately, *trehalase*, *cortex*, and *herfst* are all necessary for devel-

opment of the red autumn morph, and they appear to function at different points in a shared patterning network. Our results indicate that reaction norm variation can be both quantitative and multigenic, and we conjecture that allelic variants of these genes may act additively to shape the reaction norm curve.

Our study shows that seasonal plasticity can evolve rapidly, likely through cis-regulatory changes distributed across multiple downstream trait-specific genes. We propose that this mode of genetic evolution should be favored by selection because it allows tissue- and trait-specific tuning of reaction norms while avoiding broad effects that would otherwise be caused by changes to core seasonal response mechanisms.

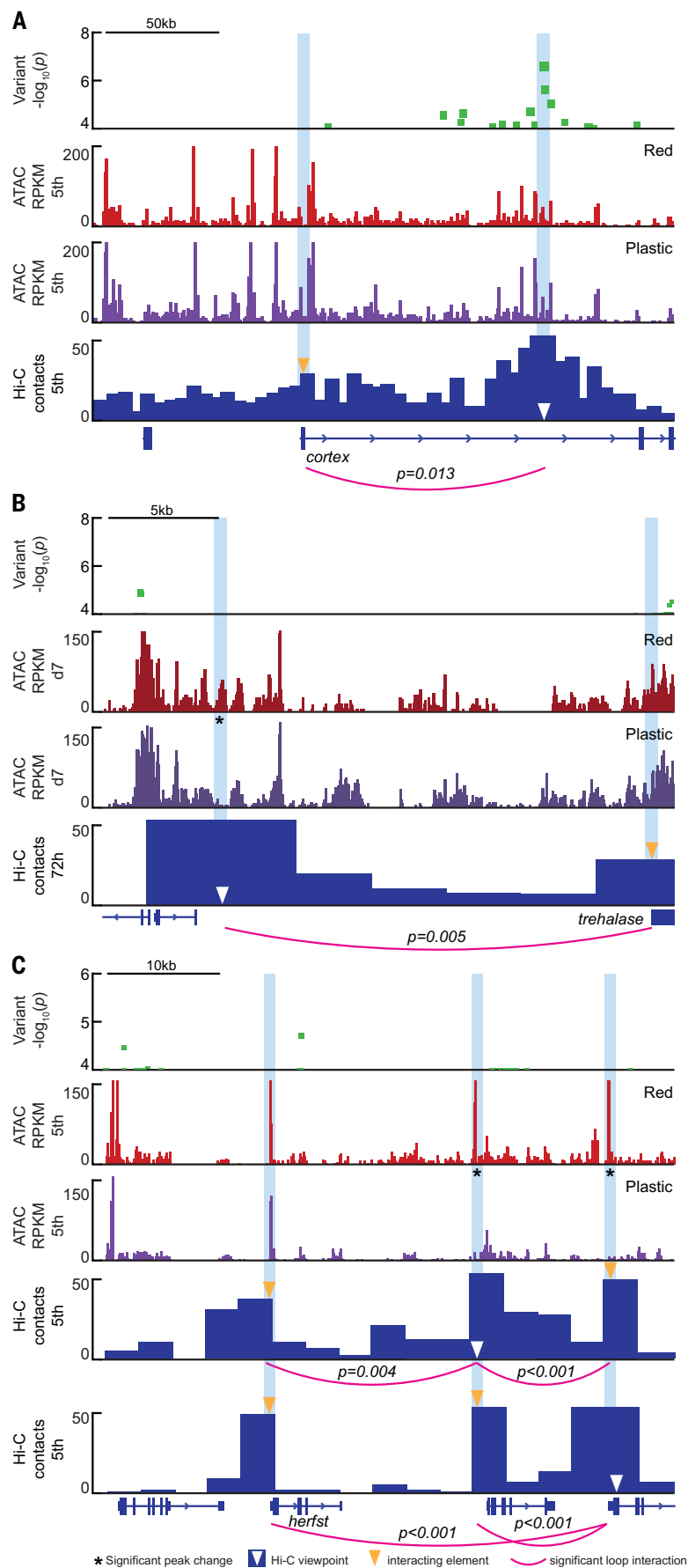


Fig. 3. Selection lines show cis-regulatory variation at causal genes. The plots show nucleotide variants associated with assimilation of red color (green), chromatin accessibility (ATAC-seq) in Red (red) and Plastic (purple) line butterflies, and Hi-C chromatin interactions (blue) using candidate CREs as viewpoints (white triangles). Lines (pink) denote significant CRE-promoter (yellow triangles) interactions, inferred from Hi-C (Fisher's exact test, $P < 0.05$). Asterisks indicate chromatin accessibility differences (Wald adjusted $P < 0.05$). **(A)** In fifth-instar wing discs, CREs under a cluster of significant nucleotide variants interact with the *cortex* promoter. **(B)** In day 7 (d7) pupae, during pigment synthesis, a CRE showing differential accessibility between lines interacts with the *trehalase* promoter. **(C)** In fifth-instar wing discs, there is a three-way chromatin interaction that includes the *herfst* promoter and two CREs that show accessibility differences between lines.

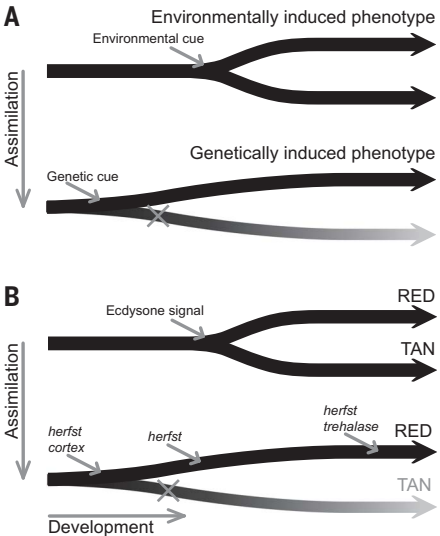


Fig. 4. Multigenic evolution of genetic assimilation. **(A)** Theoretical model where assimilation occurs by the appearance of a genetic cue that replaces the environmental cue to induce a phenotype. **(B)** In *J. coenia*, differential expression of three genes across wing development, independent of endocrine signaling, can underlie the genetic cue through which assimilation of wing color evolves.

REFERENCES AND NOTES

1. D. W. Pfennig *et al.*, *Trends Ecol. Evol.* **25**, 459–467 (2010).
2. R. F. Schneider, A. Meyer, *Mol. Ecol.* **26**, 330–350 (2017).
3. M. J. West-Eberhard, *Developmental Plasticity and Evolution* (Oxford Univ. Press, ed. 1, 2003).
4. C. H. Waddington, *Evolution* **7**, 118–126 (1953).
5. J.-M. Gibert, *Dev. Genes Evol.* **227**, 297–307 (2017).
6. K. C. Smith, *J. Res. Lepid.* **30**, 225–236 (1991).
7. D. B. Rountree, H. F. Nijhout, *J. Insect Physiol.* **41**, 1141–1145 (1995).
8. E. V. Daniels, K. A. Mooney, R. D. Reed, *Ecol. Entomol.* **37**, 155–159 (2012).
9. C. H. Waddington, *Genet. Res.* **1**, 140–150 (1960).
10. H. F. Nijhout, *Evol. Dev.* **5**, 9–18 (2003).
11. D. B. Rountree, H. F. Nijhout, *J. Insect Physiol.* **41**, 987–992 (1995).
12. N. J. Nadeau *et al.*, *Nature* **534**, 106–110 (2016).
13. A. E. Van't Hof *et al.*, *Nature* **534**, 102–105 (2016).

14. F. Figon, J. Casas, *Biol. Rev. Camb. Philos. Soc.* **94**, 156–183 (2018).
15. I. M. Ehrenreich, D. W. Pfennig, *Ann. Bot.* **117**, 769–779 (2016).
16. J.-M. Belton *et al.*, *Methods* **58**, 268–276 (2012).
17. E. V. Daniels, R. Murad, A. Mortazavi, R. D. Reed, *Mol. Ecol.* **23**, 6123–6134 (2014).

ACKNOWLEDGMENTS

We thank B. Lazarro, A. McCune, and two anonymous reviewers for helpful comments on the manuscript, and J. Kwan, S. Molyneaux, M. Kwasnica, W. Johnsen, L. Echavez, and V. Järvi for help with

butterfly husbandry. **Funding:** Supported by NSF grants IOS 1557443 and IOS 1753559 (R.D.R.). **Author contributions:** K.R.L.v.d.B. and R.D.R. conceived of and coordinated the project, acquired funding, and wrote the manuscript. K.R.L.v.d.B. and J.J.L. analyzed and curated data. K.R.L.v.d.B., J.J.L., B.J.B., R.A.F., and A.M.-V. performed experiments and collected data. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** mRNA-seq and ATAC-seq data have been submitted to GEO, #GSE135523. Additional supporting material (i.e., scripts, raw datasets, etc.) are available on Dryad: <https://doi.org/10.5061/dryad.rjdfn2z89>.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/370/6517/721/suppl/DC1
Materials and Methods
Tables S1 to S12
Figs. S1 to S9
References (18–39)

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

28 August 2019; resubmitted 16 June 2020

Accepted 1 October 2020

10.1126/science.aaz3017

CORONAVIRUS

Free fatty acid binding pocket in the locked structure of SARS-CoV-2 spike protein

Christine Toelzer^{1,2*}, Kapil Gupta^{1,2*}, Sathish K. N. Yadav^{1,2*}, Ufuk Borucu^{1,2*}, Andrew D. Davidson³, Maia Kavanagh Williamson³, Deborah K. Shoemark^{1,2}, Frederic Garzoni⁴, Oskar Stauffer^{5,6,7,8}, Rachel Milligan³, Julien Capin^{1,2}, Adrian J. Mulholland⁹, Joachim Spatz^{5,6,7,8}, Daniel Fitzgerald¹⁰, Imre Berger^{1,2,8,9†}, Christiane Schaffitzel^{1,2†}

Coronavirus disease 2019 (COVID-19), caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), represents a global crisis. Key to SARS-CoV-2 therapeutic development is unraveling the mechanisms that drive high infectivity, broad tissue tropism, and severe pathology. Our 2.85-ångstrom cryo-electron microscopy structure of SARS-CoV-2 spike (S) glycoprotein reveals that the receptor binding domains tightly bind the essential free fatty acid linoleic acid (LA) in three composite binding pockets. A similar pocket also appears to be present in the highly pathogenic severe acute respiratory syndrome coronavirus (SARS-CoV) and Middle East respiratory syndrome coronavirus (MERS-CoV). LA binding stabilizes a locked S conformation, resulting in reduced angiotensin-converting enzyme 2 (ACE2) interaction *in vitro*. In human cells, LA supplementation synergizes with the COVID-19 drug remdesivir, suppressing SARS-CoV-2 replication. Our structure directly links LA and S, setting the stage for intervention strategies that target LA binding by SARS-CoV-2.

Seven coronaviruses are known to infect humans. The four endemic human coronaviruses (HCoVs)—OC43, 229E, HKU1, and NL63—cause mild upper respiratory tract infections, whereas pandemic virus SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) and earlier SARS-CoV (severe acute respiratory syndrome coronavirus) and MERS-CoV (Middle East respiratory syndrome coronavirus) can cause severe pneumonia with acute respiratory distress syndrome, multi-organ failure, and death (1, 2).

SARS-CoV-2 has acquired functions that promote its harsh disease phenotype. SARS-CoV-2 causes severe inflammation and damage to endothelial cells in the heart, kidneys, liver, and intestines, suggestive of a vascular infection rather than a purely respiratory disease (3, 4). The attachment of SARS-CoV-2 to a host cell is initiated by the spike (S) protein trimer, which adorns the outer surface of the virus, binding to its cognate receptor angiotensin-converting enzyme 2 (ACE2) with higher affinity

than the SARS-CoV S protein (5–7). A S1-S2 polybasic furin protease cleavage site distinguishes SARS-CoV-2 from SARS-CoV and other closely related bat coronaviruses and serves to stimulate entry into host cells and cell-cell fusion (5, 8, 9). Inside the host cell, HCoVs remodel the lipid metabolism to facilitate virus replication (10). Infection by SARS-CoV-2 triggers an unusually impaired and dysregulated immune response (11) and a heightened inflammatory response (12), working in synergy with interferon production in the vicinity of infected cells to drive a feed-forward loop to up-regulate ACE2 and further escalate infection (13).

In the search for additional functions that contribute to the pathology of infection, we determined the structure of the SARS-CoV-2 S glycoprotein by cryo-electron microscopy (cryo-EM) (Fig. 1). We produced SARS-CoV-2 S as a secreted trimer (14) in MultiBac (15) baculovirus-infected Hi5 insect cells (fig. S1) (16). Highly purified protein was used for cryo-EM data collection (fig. S2 and table S1). After three-dimensional (3D) classification and refinement without applying symmetry (C1), we obtained a 3.0-Å closed conformation from 136,405 particles and a 3.5-Å open conformation with one receptor binding domain (RBD) in the up position from 57,990 particles (figs. S2 and S3). C3 symmetry was applied to the closed conformation particle pool, yielding a 2.85-Å map (Fig. 1A and figs. S2 and S3).

The structure of S displays the characteristic overall shape observed for coronavirus S proteins in the closed and open conformations (17–19), with the closed form (~70%) predominating in our dataset (Fig. 1A and figs. S2 to S4). Model building of the closed form revealed additional density in the RBDs in our structure (Fig. 1B). The tubelike shape

of this density was consistent with a fatty acid, with size and shape similar to that of linoleic acid (LA) bound to other proteins (Fig. 1B and fig. S5) (20, 21). Liquid chromatography-coupled electrospray ionization time-of-flight mass spectrometry (LC-MS) analysis confirmed the presence of a compound with the molecular weight of LA in our purified sample (Fig. 1C).

The hallmarks of free fatty acid (FFA) binding pockets in proteins are an extended “greasy” tube lined by hydrophobic amino acids, which accommodates the hydrocarbon tail, and a hydrophilic, often positively charged anchor for the acidic headgroup of the FFA. In our structure, a hydrophobic pocket mostly shaped by phenylalanines forms a bent tube into which the LA fits well (Fig. 1, D and E). The anchor for the headgroup carboxyl is provided by an arginine (Arg⁴⁰⁸) and a glutamine (Gln⁴⁰⁹) from the adjacent RBD in the trimer, giving rise to a composite LA binding site (Fig. 1E). We confirmed the presence of LA in all three binding pockets in the S trimer in the unsymmetrized (C1) closed structure (fig. S6). Similarly, masked 3D classification focusing on the RBDs could not identify any unoccupied pockets (fig. S7).

Our S construct contains alterations relative to native SARS-CoV-2 S—namely, addition of a trimerization domain and deletion of the polybasic cleavage site, neither of which alters the S conformation appreciably (14, 17) (fig. S8). Glycosylation sites are located away from the LA binding pocket and are largely native in our structure (7, 17) (table S2). Thus, neither glycosylation nor mutations are likely to affect the LA binding pocket. We compared S and RBD produced in insect cells with mammalian-produced S to identify any potential influence of differences in glycosylation on ACE2 binding by competition enzyme-linked immunosorbent assay (ELISA) (Fig. 2A). All three reagents bound ACE2 efficiently. We then used size exclusion chromatography (SEC) with purified proteins to further confirm ACE2 binding by S (Fig. 2B). The LA binding pocket and the receptor binding motif (RBM) are distal and nonoverlapping (Fig. 2C). Notably, in the LA-bound S, the RBM is ordered and buried at the interface between RBDs, whereas it was disordered in previously described SARS-CoV-2 S cryo-EM structures (7, 17).

SARS-CoV-2 S can also adopt an open conformation (fig. S4), which is compatible with binding ACE2. In previous apo S cryo-EM structures, about 60 to 75% of the S trimers were in the open conformation (7, 17), in contrast to our observation of ~70% in the closed conformation. This result could be due to LA stabilizing the closed conformation; if so, LA would be expected to reduce ACE2 binding. We performed surface plasmon resonance (SPR) experiments with biotinylated ACE2 immobilized on a streptavidin-coated chip

¹School of Biochemistry, University of Bristol, 1 Tankard's Close, Bristol BS8 1TD, UK. ²Bristol Synthetic Biology Centre, BrisSynBio, 24 Tyndall Ave., Bristol BS8 1TQ, UK. ³School of Cellular and Molecular Medicine, University of Bristol, University Walk, Bristol BS8 1TD, UK. ⁴Imophoron Ltd., St. Philips Central, Albert Rd., St. Philips, Bristol BS2 0XJ, UK. ⁵Department for Cellular Biophysics, Max Planck Institute for Medical Research, Jahnstraße 29, 69120 Heidelberg, Germany. ⁶Institute for Physical Chemistry, Department for Biophysical Chemistry, University of Heidelberg, Im Neuenheimer Feld 253, 69120 Heidelberg, Germany. ⁷Max Planck School Matter to Life, Jahnstraße 29, D-69120 Heidelberg, Germany. ⁸Max Planck Bristol Centre for Minimal Biology, Cantock's Close, Bristol BS8 1TS, UK. ⁹School of Chemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, UK. ¹⁰Geneva Biotech Sàrl, Avenue de la Roseraie 64, 1205, Geneva, Switzerland.

*These authors contributed equally to this work.

†Corresponding author. Email: cb14941@bristol.ac.uk (C.S.); imre.berger@bristol.ac.uk (I.B.)

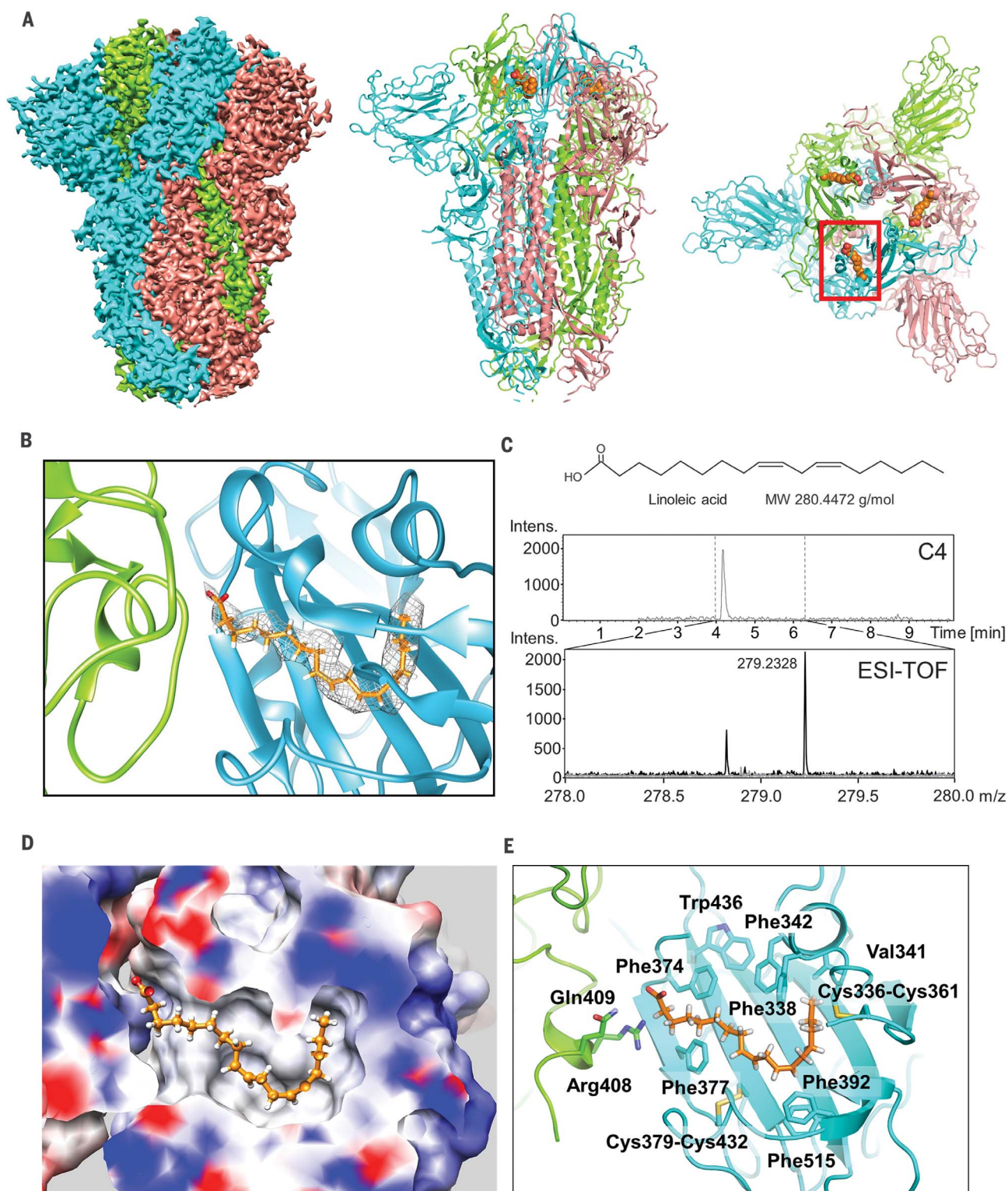


Fig. 1. Cryo-EM structure of the SARS-CoV-2 S linoleic acid complex.

(A) Cryo-EM density of the S trimer is shown (left). Monomers are in cyan, green, and pink, respectively. The structure is also shown in a cartoon representation in a front view (middle) and top view (right). Bound LA is illustrated as orange spheres. One LA binding pocket is surrounded by the red box. (B) Composite LA binding pocket formed by adjacent RBDs. Tube-shaped EM density is shown. (C) LC-MS analysis of purified S. (Top) Chemical structure and molecular weight (MW) of LA. (Middle) C4 column elution

profile. (Bottom) Electrospray ionization time-of-flight (ESI-TOF) spectra of wash solution (gray) and C4 peak elution fraction (black), with peak molecular weight indicated. Intens., intensity, m/z, mass/charge ratio. (D) Hydrophobic LA binding pocket in a surface representation, illustrating the excellent fit of bound LA (orange; shown in ball-and-stick representation). Blue and red indicate positive and negative surface charge, respectively. (E) LA interactions with amino acids in the binding pocket. The acidic LA headgroup is in the vicinity of an arginine (Arg⁴⁰⁸) and a glutamine (Gln⁴⁰⁹).

Fig. 2. Functional characterization of LA-bound SARS-CoV-2 S.

(A) Insect cell (Hi5)-expressed S (dark blue bars), insect cell-expressed RBD (light blue bars), and mammalian (HEK293)-expressed S (white bars) in competition ELISAs with immobilized ACE2. Error bars indicate SDs (three replicates).

(B) Interaction of LA-bound SARS-CoV-2 S protein with ACE2 was analyzed by SEC, evidencing complex formation. (Left) SEC profiles are shown for ACE2 (yellow, III), LA-bound S (green, II), and a mixture of ACE2 and LA-bound S (orange, I). (Right) Peak fractions (I to III) were analyzed by SDS-polyacrylamide gel electrophoresis, evidencing the expected proteins.

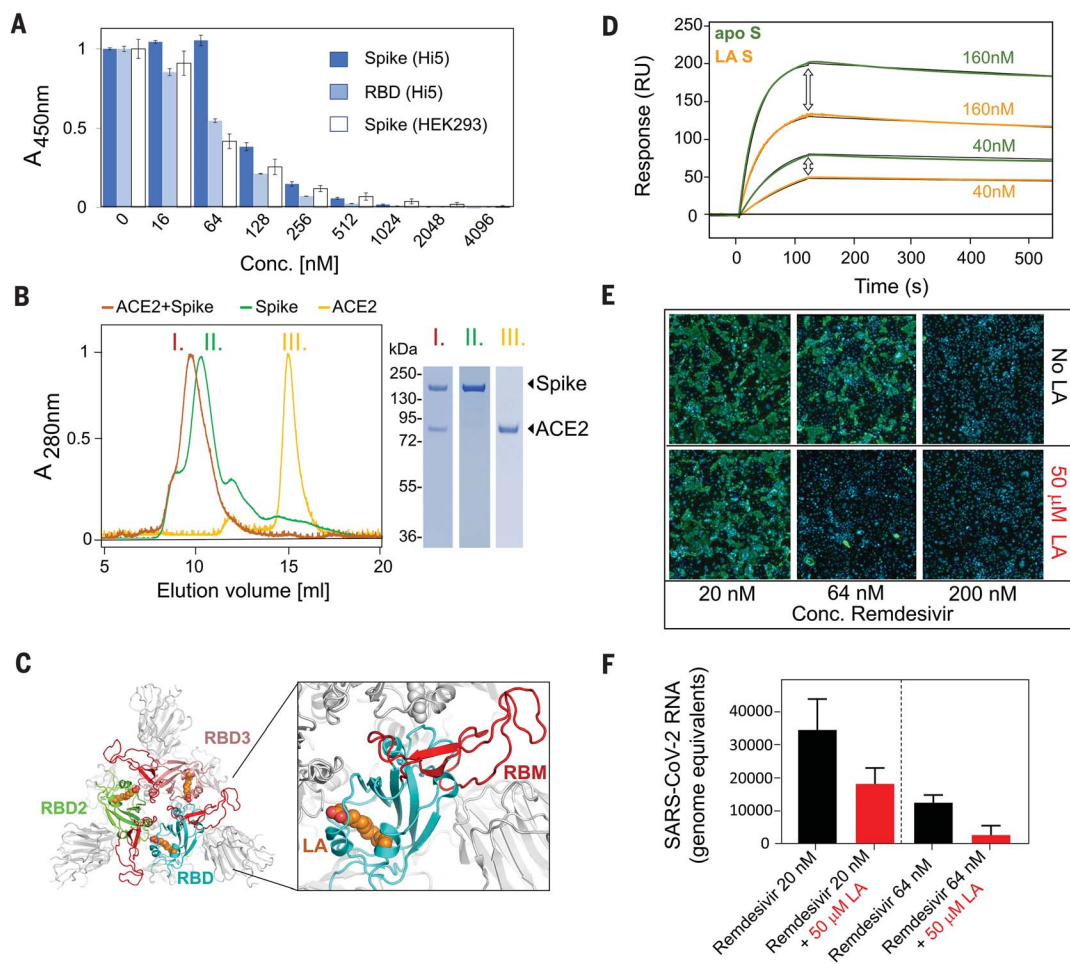
(C) (Left) Top view of the LA-bound S glycoprotein trimer, with RBDs shown in cyan, green, and pink. In each RBD subunit, the motif responsible for ACE2 binding (RBM) is in red, and LA is shown as orange spheres. (Right) A close-up view into the cyan RBD shows that the RBM is fully ordered and that LA and RBM are not in direct contact.

(D) SPR analysis of binding of the LA-bound S trimer (orange curves) and the apo S trimer (green curves) to immobilized ACE2. Apo S and LA-bound S were diluted to concentrations of 40 and 160 nM, respectively. Black lines correspond to a global fit from a 1:1 binding model. RU, resonance units.

(E) Synergistic effect of LA and remdesivir on SARS-CoV-2 replication. Effects of varying doses of remdesivir \pm 50 μ M LA on virus infection are shown. Human Caco-2 ACE2+ cells were infected with SARS-CoV-2 and then treated with varying doses of

remdesivir \pm 50 μ M LA. At 96 hours after infection, cells were fixed and infected cells were detected by immunofluorescence assay using an anti-N antibody (green). Cell nuclei were stained by 4',6-diamidino-2-phenylindole (DAPI; blue). Representative images corresponding to the remdesivir dose range of 20 to 200 nM are shown.

(F) Amount of extracellular virus present in wells ($n = 3$) at the dose combinations shown was determined by quantitative reverse transcription polymerase chain reaction (performed in duplicate for each sample). Error bars show SD.



(Fig. 2D and fig. S9). We first determined the dissociation constant (K_D) of the RBD-ACE2 interaction to validate our assay. Our value (26 nM; fig. S9C) is in good agreement with that of previous work [44 nM (22)], which used SPR with the RBD immobilized and ACE2 as an analyte. Apo S was prepared by applying Lipidex, the established method for removing lipids from lipid-binding proteins (23). A K_D of 0.7 nM was obtained for the apo S-ACE2 interaction (fig. S9A). For the LA-bound S-ACE2 interaction, we obtained a K_D of 1.4 nM (fig. S9B). We consistently obtained a markedly reduced resonance unit signal for LA-bound S compared with apo S at the same concentrations (Fig. 2D and fig. S9, A and B). This observation is consistent with the apo state having a higher percentage of S trimers in the open, ACE2-accessible conformation.

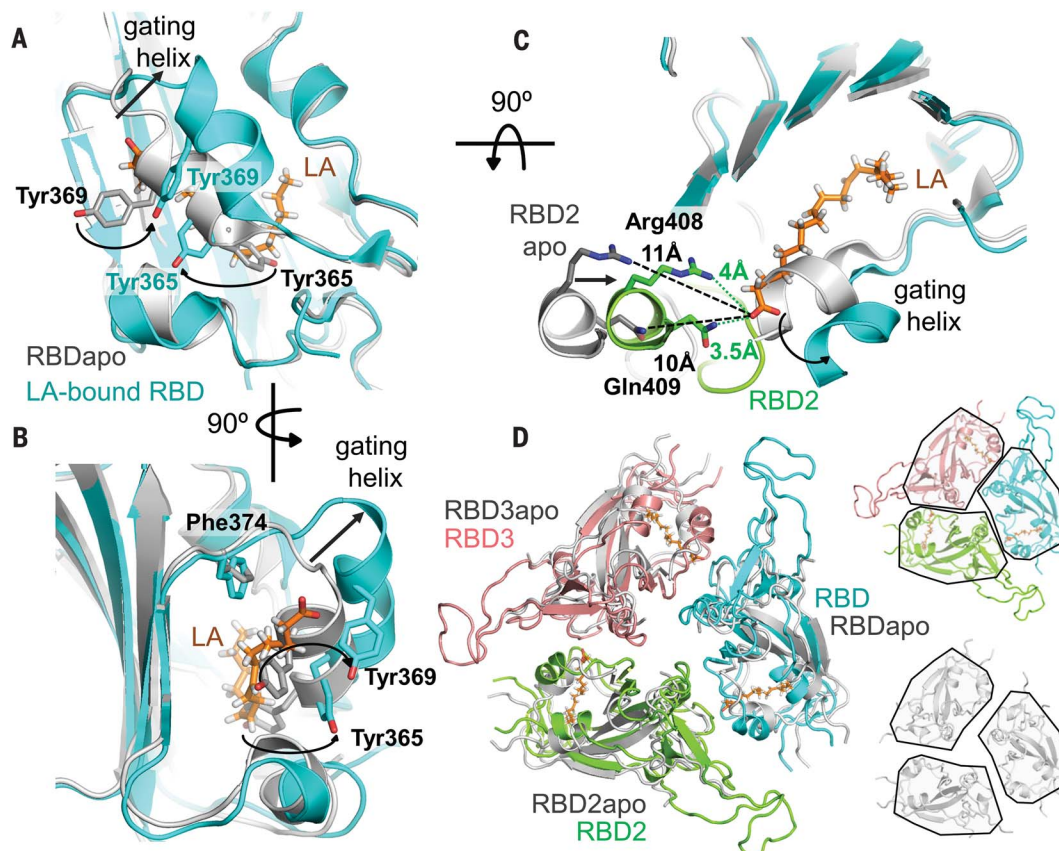
We characterized the affinity of the LA interaction both experimentally and computationally. Our SPR assays with immobilized RBD yielded a binding constant of \sim 41 nM, exhibiting a slow off-rate, consistent with tight binding of LA (fig. S10). Repeated molecular dynamics simulations of the entire locked LA-bound S trimer (three 100-ns simulations) using GROMACS-2019 (24) corroborated the persistence of stable interactions between LA and the S trimer (movies S1 and S2). The affinity of LA binding to the S trimer will likely be higher than that of binding to the RBD alone, taking into account polar headgroup interactions with Arg⁴⁰⁸ and Gln⁴⁰⁹ of the adjacent RBD (Fig. 1E). The resolution of the RBDs in our open S cryo-EM structure was insufficient to either assign or rule out a ligand-bound pocket (fig. S3). However, the

slow off-rate observed with the RBD monomer (fig. S10) suggests that LA binding could be maintained when the S trimer transiently converts into the open conformation. This is supported by our observation that LA was retained during S purification, in spite of S trimers adopting the open form \sim 30% of the time (fig. S2), and by our MD simulations with a modeled ligand-bound open S trimer (movie S3) in which all three LAs remained bound for 500 ns.

Next, we investigated the effect of LA in experiments with live SARS-CoV-2 infecting human epithelial cells. Remdesivir is an RNA-dependent RNA polymerase inhibitor and the first antiviral drug to show a benefit in the treatment of coronavirus disease 2019 (COVID-19) in clinical trials, albeit with considerable side effects at the doses required (25). LA supplementation at concentrations of 50 to 100 μ M

Fig. 3. Comparison of LA-bound and apo S structures.

(A) Superimposition of LA-bound SARS-CoV-2 RBD (cyan) and ligand-free apo RBD (gray) [PDB ID 6VXX (7)]. The gating helix at the entrance of the hydrophobic pocket moves by 6 Å in the presence of LA. Tyr³⁶⁵ and Tyr³⁶⁹ swing away, avoiding clashes with LA (orange). Black arrows indicate the rearrangements. (B) Same structure as in (A) rotated by 90° as indicated, showing the entrance of the hydrophobic pocket. (C) Formation of a composite LA binding pocket by two adjacent RBDs in LA-bound S involves a ~5-Å movement of RBD2 (green) toward RBD1 (cyan) as opposed to apo S (gray). (D) Superimposition of the RBD trimer of apo S (gray) and LA-bound S (RBD1, cyan; RBD2, green; RBD3, pink; LA, orange) is shown (left). The individual RBD trimers are depicted for LA-bound S (right, top) and apo S (right, bottom), with RBDs boxed in black, highlighting the compaction of RBDs in the LA-bound S structure.



was previously shown to affect coronavirus entry and replication (10). We administered remdesivir at concentrations of 20, 64, and 200 nM, supplementing with 50 µM LA (Fig. 2E). Our results revealed synergy, with the dose of remdesivir required to suppress SARS-CoV-2 replication markedly reduced by the addition of LA (Fig. 2, E and F).

We superimposed our LA-bound structure on previous SARS-CoV-2 apo S structures in the closed conformation (7, 17) and identified a gating helix located directly at the entrance of the binding pocket (Fig. 3, A to C). This gating helix, which comprises Tyr³⁶⁵ and Tyr³⁶⁹, is displaced by ~6 Å when LA is bound, thus opening the pocket (Fig. 3, A and B). In the apo SARS-CoV-2 S trimer (7, 17), a gap between adjacent RBDs places the hydrophilic anchor residues ~10 Å from the position of the LA headgroup (Fig. 3C). Upon LA binding, the adjacent RBD in the trimer moves toward its neighbor, and the anchor residues Arg⁴⁰⁸ and Gln⁴⁰⁹ lock down on the headgroup of LA (Fig. 3, C and D). Overall, this results in a compaction of trimer architecture in the region formed by the three RBDs, producing a locked S structure (Fig. 3D and movie S4).

We investigated whether the LA-binding pocket is conserved in the seven coronaviruses

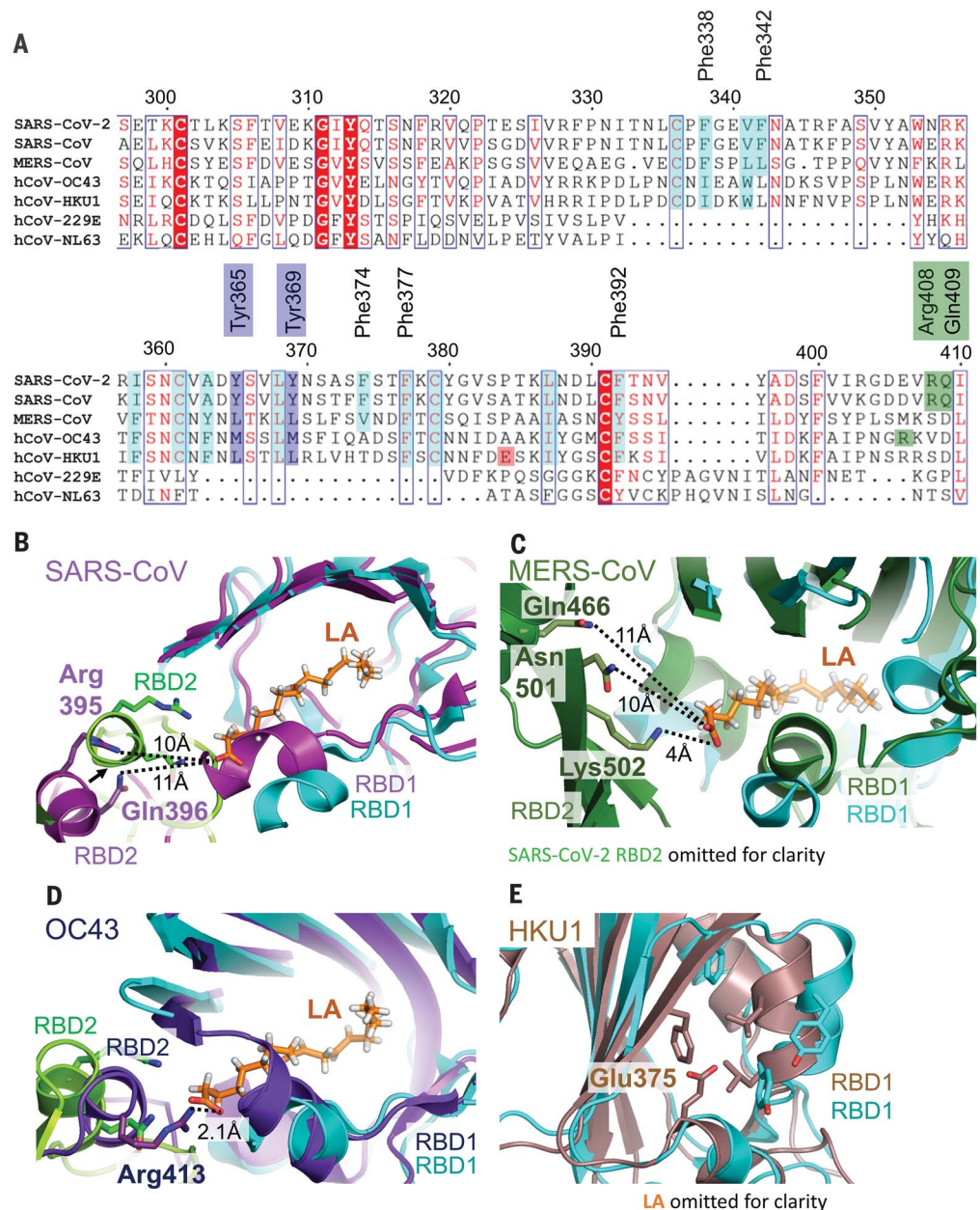
that infect humans (Fig. 4A and table S3). Sequence alignment shows that all residues lining the hydrophobic pocket and the anchor residues (Arg⁴⁰⁸ and Gln⁴⁰⁹) in SARS-CoV-2 are fully conserved in SARS-CoV (Fig. 4A). Structural alignment of LA-bound RBDs within the trimer of SARS-CoV-2 and apo SARS-CoV RBDs (19) reveals that the LA binding pocket is present in SARS-CoV. The greasy tube is flanked by a gating helix as in SARS-CoV-2, with Arg³⁹⁵ and Gln³⁹⁶ of SARS-CoV positioned 10 and 11 Å from the entrance, respectively, in a conformation that is virtually identical to that of apo SARS-CoV-2 (Figs. 3C and 4B). In MERS-CoV, the gating helix and hydrophobic residues lining the pocket are also present. Tyr³⁶⁵, Tyr³⁶⁹, and Phe³⁷⁴ are substituted by likewise hydrophobic leucines and a valine, respectively (Fig. 4, A and C) (19). The Arg⁴⁰⁸-Gln⁴⁰⁹ pair is not conserved; however, we identify Asn⁵⁰¹, Lys⁵⁰², and Gln⁴⁶⁶ as potential anchor residues, located on a β sheet and an α helix within the adjacent RBD, up to 11 Å from the entrance (Fig. 4C). Thus, the greasy tube and hydrophilic anchor appear to be present in MERS-CoV, suggesting convergent evolution. In HCoV OC43, the gating helix and hydrophobic residues that line the pocket are largely conserved, whereas Tyr³⁶⁵, Tyr³⁶⁹, and Phe³⁷⁴ are

replaced by methionines and alanine, respectively (Fig. 4A) (18). Arg⁴¹³ is located on the same helix as Arg⁴⁰⁸ and Gln⁴⁰⁹ in SARS-CoV-2 and could serve as a hydrophilic anchor (Fig. 4D). No gap exists in this presumed apo form structure between the RBDs, which appear already in the locked conformation (Fig. 4D and fig. S11) (18). In HCoV HKU1, the hydrophobic residues are again largely conserved, but a charged residue (Glu³⁷⁵) is positioned directly in front of the entrance, obstructing access for a putative ligand (Fig. 4E) (26). The RBDs of HCoVs 229E and NL63 adopt a very different fold (fig. S13) (27, 28), and many of the LA binding residues are not present (Fig. 4A), thus hampering predictions of a binding site for fatty acids.

In summary, we find four molecular features that mediate LA binding to SARS-CoV-2 and potentially also to SARS-CoV and MERS-CoV S proteins: a conserved hydrophobic pocket, a gating helix, amino acid residues prepositioned to interact with the LA carboxyl headgroup, and loosely packed RBDs in the apo form. By contrast, in each of the four common circulating HCoVs, it appears that one or more of these four architectural prerequisites is lacking in the S protein structures (Fig. 4 and figs. S11 and S12). LA binding to SARS-CoV-2 S triggers a locking

Fig. 4. Human coronavirus RBD architectures.

(A) Alignments of the seven CoV strains that can infect humans, highlighting conserved residues. Cyan, residues lining the hydrophobic pocket; purple, gating helix residues; green, residues positioned to interact with the LA polar headgroup; and red, Glu³⁷⁵ in HKU1 [see panel (E)]. 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. **(B)** Superimposition of RBD1 of LA-bound SARS-CoV-2 (RBD2, green) with RBD1 of ligand-free apo SARS-CoV [RBD1 and RBD2, magenta; PDB ID 5X58 (19)] indicates a conservation of the composite binding pocket. **(C)** Superimposition of RBD1 of LA-bound SARS-CoV-2 (RBD2 is omitted for clarity) with RBD1 of MERS-CoV [RBD1 and RBD2, forest green; PDB ID 5X5F (19)]. **(D)** Superimposition of RBD1 of LA-bound SARS-CoV-2 (RBD2, green) with RBD1 of OC43 [RBD1 and RBD2, purple; PDB ID 6NZK (18)]. **(E)** Superimposition of LA-bound SARS-CoV-2 RBD with HKU1 RBD [brown; PDBID 5GNB (26)]. LA is omitted in SARS-CoV-2 RBD for clarity.



down of the hydrophilic anchor and a compaction of the RBD trimer (Fig. 3, C and D). In addition to stabilizing the closed conformation, this lockdown could also help stabilize the S1 region, which includes the N-terminal domain and the RBD. The RBM, central to ACE2 binding, appears to be conformationally preorganized in our structure (Fig. 2C), indicating a generally more rigid RBD trimer when LA is bound. Although direct cross-talk between the LA binding pocket and the RBM is not apparent from our structure (Fig. 2C), the conformational changes in the RBD trimer (Fig. 3) could affect ACE2 docking and infectivity, as indicated by our SPR assays that show reduced levels of S binding in the presence of LA (Fig. 2D). The S protein's tight

binding of LA originates from a well-defined size and shape complementarity afforded by the pocket (Fig. 1, B and D). The LA binding pocket thus presents a promising target for future development of small-molecule inhibitors that, for example, could irreversibly lock S in the closed conformation and interfere with receptor interactions. It is noteworthy in this context that a fatty acid binding pocket was exploited previously to develop potent small-molecule antiviral drugs to treat rhinovirus, locking viral surface proteins in a conformation incompatible with receptor binding (29, 30). These antivirals were successful in human clinical trials (31, 32).

A recent proteomic and metabolomic study of COVID-19 patient sera showed continuous

decrease of FFAs, including LA (33). Lipid metabolome remodeling is a common element of viral infection (34, 35). For coronaviruses, the LA-to-arachidonic acid metabolic pathway was identified as central to lipid remodeling (10). We hypothesize that LA sequestration by SARS-CoV-2 could confer a tissue-independent mechanism by which pathogenic coronavirus infection may drive immune dysregulation and inflammation (35–37). Our findings provide a direct structural link between LA, COVID-19 pathology, and the virus itself and suggest that both the LA binding pocket within the S protein and the multinodal LA signaling axis represent excellent therapeutic intervention points to treat SARS-CoV-2 infections.

REFERENCES AND NOTES

- P. Zhou et al., *Nature* **579**, 270–273 (2020).
- A. Zumla, J. F. Chan, E. I. Azhar, D. S. Hui, K. Y. Yuen, *Nat. Rev. Drug Discov.* **15**, 327–347 (2016).
- Z. Varga et al., *Lancet* **395**, 1417–1418 (2020).
- V. G. Puelles et al., *N. Engl. J. Med.* **383**, 590–592 (2020).
- M. Hoffmann et al., *Cell* **181**, 271–280.e8 (2020).
- M. Letko, A. Marzi, V. Munster, *Nat. Microbiol.* **5**, 562–569 (2020).
- A. C. Walls et al., *Cell* **181**, 281–292.e6 (2020).
- M. Hoffmann, H. Kleine-Weber, S. Pöhlmann, *Mol. Cell* **78**, 779–784.e5 (2020).
- S. Matsuyama et al., *Proc. Natl. Acad. Sci. U.S.A.* **117**, 7001–7003 (2020).
- B. Yan et al., *Viruses* **11**, 73 (2019).
- C. Qin et al., *Clin. Infect. Dis.* **71**, 762–768 (2020).
- M. Z. Tay, C. M. Poh, L. Rénia, P. A. MacAry, L. F. P. Ng, *Nat. Rev. Immunol.* **20**, 363–374 (2020).
- C. G. K. Ziegler et al., *Cell* **181**, 1016–1035.e19 (2020).
- F. Amanat et al., *Nat. Med.* **26**, 1033–1036 (2020).
- D. J. Fitzgerald et al., *Nat. Methods* **3**, 1021–1032 (2006).
- To produce S glycoprotein, we used baculovirus-infected insect cells (Hi5) cultured in ESF 921 media from Expression Systems, which contains cod liver oil as a nutrient supplement. Cod liver oil contains hundreds of FFAs, including LA (38), a possible source of the fatty acid. In parallel to our work, other researchers determined cryo-EM structures of S expressed in a mammalian system (HEK293) in serum and protein-free media (39, 40). The tube-shaped density we identified in our study is also present in those structures (but was not assigned or interpreted). Thus, we conclude that LA binding in the SARS-CoV-2 S pocket is not dependent on the expression system or media used.
- D. Wrapp et al., *Science* **367**, 1260–1263 (2020).
- M. A. Tortorici et al., *Nat. Struct. Mol. Biol.* **26**, 481–489 (2019).
- Y. Yuan et al., *Nat. Commun.* **8**, 15092 (2017).
- J. M. Gullett, M. G. Cuyppers, M. W. Frank, S. W. White, C. O. Rock, *J. Biol. Chem.* **294**, 16416–16428 (2019).
- J. Wang, E. J. Murphy, J. C. Nix, D. N. M. Jones, *Sci. Rep.* **10**, 3300 (2020).
- J. Shang et al., *Nature* **581**, 221–224 (2020).
- J. F. Glatz, J. H. Veerkamp, *J. Biochem. Biophys. Methods* **8**, 57–61 (1983).
- H. J. C. Berendsen, D. van der Spoel, R. van Drunen, *Comput. Phys. Commun.* **91**, 43–56 (1995).
- J. H. Beigel et al., *N. Engl. J. Med.* 10.1056/NEJMoa2007764 (2020).
- X. Ou et al., *Nat. Commun.* **8**, 15216 (2017).
- Z. Li et al., *eLife* **8**, e51230 (2019).
- A. C. Walls et al., *Nat. Struct. Mol. Biol.* **23**, 899–905 (2016).
- J. Badger et al., *Proc. Natl. Acad. Sci. U.S.A.* **85**, 3304–3308 (1988).
- M. A. Oliveira et al., *Structure* **1**, 51–68 (1993).
- V. Casanova, F. H. Sousa, C. Stevens, P. G. Barlow, *Future Virol.* **13**, 505–518 (2018).
- K. K. W. To, C. C. Y. Yip, K. Y. Yuen, *J. Formos. Med. Assoc.* **116**, 496–504 (2017).
- B. Shen et al., *Cell* **182**, 59–72.e15 (2020).
- C. M. Goodwin, S. Xu, J. Munger, *Trends Microbiol.* **23**, 789–798 (2015).
- J. W. Schoggins, G. Randall, *Cell Host Microbe* **14**, 379–385 (2013).
- M. M. Zaman et al., *Am. J. Physiol. Lung Cell. Mol. Physiol.* **299**, L599–L606 (2010).
- I. Kimura, A. Ichimura, R. Ohue-Kitano, M. Igarashi, *Physiol. Rev.* **100**, 171–210 (2020).
- S. Hauff, W. Vetter, *J. Agric. Food Chem.* **57**, 3423–3430 (2009).
- Y. Cai et al., *Science* **369**, 1586–1592 (2020).
- A. G. Wrobel et al., *Nat. Struct. Mol. Biol.* **27**, 763–767 (2020).

ACKNOWLEDGMENTS

We thank all members of the Berger and Schaffitzel teams as well as G. Singh, Y. Yamauchi, and D. Matthews (University of Bristol, UK) for assistance; F. Krammer (Icahn School of Medicine, USA) for sharing expression plasmids; V. Chang and R. Aricescu (MRC-LMB, UK) for the gift of S expressed in HEK293; A. Finn (Bristol UNCOVER Group and Children's Vaccine Centre, Bristol Medical School), J. Tavaré (School of Biochemistry, Bristol), K. Gillespie (Diabetes and Metabolism Unit, Southmead Hospital, Univ. of Bristol), and D. Fitzgerald (Quest Imaging Medical Associates, USA) for helpful discussions and careful reading of the manuscript; S. Burbidge, T. Batstone, and M. Williams for computation infrastructure support; the Advanced Computing Research Centre (ACRC) at the University of Bristol for access to BlueCryo, BlueCrystal Phase 4, and BlueGEM; and the UK HECBioSim for access to the UK supercomputer, ARCHER. We are particularly grateful to T. Thangarajah (Genscript) for early access to Genscript's cPass SARS-CoV-2 Neutralization Antibody Detection/Surrogate Virus Neutralization Test Kit (L00847). We also thank S. Fabritz and the Core Facility for Mass Spectrometry at the Max Planck Institute for Medical Research for their support on MS measurements. **Funding:** This research received support from the Elizabeth Blackwell Institute for Health Research and the EPSRC Impact Acceleration Account EP/R511663/1, University of Bristol; from BrisSynBio, a BBSRC/EPSC Research Centre for synthetic biology at the University of Bristol (BB/L01386X/1 to I.B., A.J.M., and D.K.S.); and from the BBSRC (BB/P000940/1 to C.S. and I.B.). This work received generous support from the Oracle Higher Education and Research program to enable cryo-EM data processing using Oracle's high-performance public cloud infrastructure (https://cloud.oracle.com/en_US/cloud-infrastructure) and from the EPSRC through a COVID-19 project award via HECBioSim to access ARCHER (to A.J.M.). We acknowledge support and assistance by the Wolfson Bioimaging Facility and the GW4 Facility for High-Resolution Electron Cryo-Microscopy funded by the Wellcome Trust (202904/Z/16/Z and 206181/Z/17/Z) and BBSRC (BB/R000484/1). We would like to acknowledge support of the University of Bristol's Alumni and Friends, who funded the ImageXpress Pico Imaging System. O.S. acknowledges support from the Elisabeth Muerer Foundation, the Max Planck School Matter to Life, and the Heidelberg Biosciences International Graduate School. A.D.D. is supported by the U.S. Food

and Drug Administration (HHSF223201510104C) and the UK Research and Innovation/Medical Research Council (MRC) (MR/V027506/1). M.K.W. is supported by MRC grants MR/R020566/1 and MR/V027506/1 (awarded to A.D.D.). A.J.M. is supported the British Society for Antimicrobial Chemotherapy (BSAC-COVID-30) and the EPSRC (EP/M022609/1, CCP-BioSim). I.B. acknowledges support from the EPSRC Future Vaccine Manufacturing and Research Hub (EP/R013764/1). C.S. and I.B. are Investigators of the Wellcome Trust (210701/Z/18/Z and 106115/Z/14/Z). **Author contributions:** C.S. and I.B. conceived and guided the study. F.G., K.G., and J.C. produced, purified, and analyzed samples. K.G. carried out biochemical experiments. S.K.N.Y. and U.B. prepared grids and collected EM data. S.K.N.Y., U.B., K.G., and C.T. carried out image analysis and model building. A.D.D., M.K.W., and R.M. performed all live virus CL3 work and analyzed data. D.K.S. and A.J.M. performed all MD simulations. O.S. and J.S. performed and interpreted mass spectrometry. C.T., K.G., D.F., I.B., and C.S. interpreted results. D.F., I.B., and C.S. wrote the manuscript, with input from all authors. **Competing interests:** The authors declare competing interests. I.B. and F.G. report shareholding in Imophoron Ltd. that is unrelated to this work. I.B. and D.F. report shareholding in Geneva Biotech Sàrl that is related to this work. Patent applications describing methods and material compositions based on the present observations have been filed. **Data and materials availability:** Datasets generated during this study have been deposited in the Electron Microscopy Data Bank (EMDB) under accession numbers EMD-11145 (C3 closed conformation), EMD-11144 (C1 closed conformation), and EMD-11146 (open conformation) and in the Protein Data Bank (PDB) under accession numbers: 6ZB5 (C3 closed conformation) and 6ZB4 (C1 closed conformation). Reagents are available from I.B. and C.S. under a material transfer agreement with the University of Bristol. 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/370/6517/725/suppl/DC1
Materials and Methods
Figs. S1 to S12
Tables S1 to S3
References (41–63)
MDAR Reproducibility Checklist
Movies S1 to S4

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

15 June 2020; accepted 16 September 2020
Published online 21 September 2020
10.1126/science.abd3255

Who's the top employer for 2020?

Science Careers' annual survey reveals the top companies in biotech & pharma voted on by *Science* readers.

Read the article and employer profiles at sciencecareers.org/topemployers



Science 2020 TOP EMPLOYER





中北大学
NORTH UNIVERSITY OF CHINA

Welcome to Join Us

Brief Introduction

Sharing the only garden-like campus in Taiyuan, a beautiful historic city of over 2,500 years old in Shanxi Province, North University of China (NUC) is a multi-disciplinary teaching and research university including engineering, science, liberal arts, economics, management, law, art and education. NUC consists of 1 national key discipline, 15 provincial key advantage disciplines and 6 centers for post-doctoral studies, and is authorized by the Academic Degrees Committee of the State Council to offer 24 PhD programs in 7 first-level disciplines, 64 Master programs in 23 first-level disciplines and 85 Bachelor programs.

Up to now, NUC has achieved tremendous accomplishments in a great number of research projects, some of which have won high-level awards, including 5 Second Prizes of National Award for Science & Technology and more than 100 Provincial Awards for Science & Technology.

Recruitment

1. Postdoctoral scholar/fellow holding formal research positions in overseas universities, or Graduate student with PhD of overseas universities.
2. Distinguished talents with different titles of National-Level and Provincial-Level Scholars, Known Associated Professor of universities home and abroad and Doctors with remarkable achievements.

Payment and Welfare

Generous Settling-down Allowance, Research Start-Up Funds, On-duty Allowance and ample Housing will be offered.

Contact

Tel: 86+03513924993
Fax: 86+03513924993
Email: rsglk@nuc.edu.cn
Website: www.nuc.edu.cn

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.
- A tool for setting strategic goals for the coming year, with optional reminders to keep you on track.
- Articles and resources to guide you through the process.
- Options to save materials online and print them for further review and discussion.
- Ability to select which portion of your IDP you wish to share with advisors, mentors, or others.
- A certificate of completion for users that finish myIDP.



Visit the website and start planning today!
myIDP.sciencecareers.org

ScienceCareers In partnership with: AAAS



**SOMETIMES
THE GRASS
REALLY
IS GREENER
SOMEPLACE
NEW.**

 Find your next job at [ScienceCareers.org](https://sciencecareers.org)

ScienceCareers
FROM THE JOURNAL SCIENCE 

Faculty Cluster Hires in...

The University of Texas at San Antonio (UTSA) is recruiting world-class researchers as part of a Clustered & Connected Hiring Program. We are seeking faculty members to foster collaborative research, education, and outreach and create transdisciplinary areas of knowledge that will advance their respective fields. We invite applications for the following research cluster...

Human Performance (3 positions)

- » Department Chair in Kinesiology in the College for Health, Community and Policy
- » Assistant or Associate Professor in Data Science, a joint appointment with the College for Health, Community and Policy and the College of Engineering
- » Department Chair of Mechanical Engineering and Professor in Biomechanics in the College of Engineering

More information can be found at <https://jobs.utsa.edu>
Review of applications will begin **November 2020**.

Applications received after this date will be accepted and reviewed until a position is filled.

For further information, contact Bernard.Arulanandam@utsa.edu

As an equal employment opportunity and affirmative action employer, it is the policy of UTSA to promote and ensure equal employment opportunity for all individuals without regard to race, color, religion, sex, national origin, age, sexual orientation, disability, or veteran status. The University is committed to the Affirmative Action Program in compliance with all government requirements to ensure nondiscrimination. Women and minorities are encouraged to apply. The UTSA campus is accessible to persons with disabilities.

UTSA

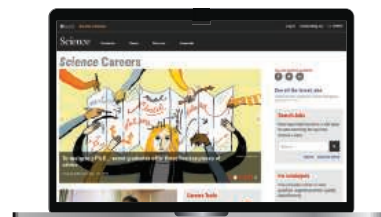
The University of Texas
at San Antonio™

CAREER PLANNING

**Science Careers helps you advance
your career. Learn how !**

- Register for a free online account on [ScienceCareers.org](https://sciencecareers.org).
- Search hundreds of job postings and find your perfect job.
- Sign up to receive e-mail alerts about job postings that match your criteria.
- Upload your resume into our database and connect with employers.
- Watch one of our many webinars on different career topics such as job searching, networking, and more.
- Download our career booklets, including Career Basics, Careers Beyond the Bench, and Developing Your Skills.
- Complete an interactive, personalized career plan at "my IDP"
- Visit our Employer Profiles to learn more about prospective employers.
- Read relevant career advice articles from our library of thousands.

Visit [ScienceCareers.org](https://sciencecareers.org)
today — all resources are free



SCIENCECAREERS.ORG

ScienceCareers
FROM THE JOURNAL SCIENCE 

By Thomas Gernon

A sabbatical reboot

“Why have I traded my private office for a desk in a shared room?” I wondered. “Will I get any work done?” It was my first day at my new co-working space on the outskirts of town, 10 kilometers from my university. “I’m here on sabbatical,” I announced to my new office-mates. They were directors of tech startups, unaccustomed to mingling with geoscientists like me. I desperately needed some breathing space away from the pressures of university life, but I wasn’t totally convinced it would work out. I needn’t have worried. The co-working space turned out to be the perfect environment for rebooting my flailing research career.

Before my sabbatical, I was on the verge of leaving academia. Six years into an assistant professorship, I felt trapped in my narrow specialty: the science of how volcanic eruptions unearth diamonds. My work had been well funded by the diamond industry when I was a grad student and postdoc. But during my first year as a professor, the 2008 financial crisis hit. Diamond sales plummeted and funding dried up. I tried to pivot by proposing projects that made use of my lab skills in other research areas, such as fluid mechanics. But those proposals were rejected, too. After a productive Ph.D., I wondered, “Did I peak too soon?”

My publication record as a faculty member was soon among the worst in my department. I wasn’t worried about being denied tenure and losing my job because my university doesn’t have a tenure system. But without a vibrant research career, I became increasingly restless. Something big had to change. Then I got thrown a lifeline: My university approved my request for a semester-long sabbatical, which would give me the mental space to think carefully about how to move forward.

At first, I found the hustle and bustle of the co-working space—an open-plan office I shared with eight other people—disruptive. I also missed having geoscientists around to bounce ideas off. Very soon, though, I settled in. I bought noise canceling headphones so I could focus when I needed to. And I found the fast-moving, entrepreneurial environment exhilarating. My co-workers were curious about the problems I was addressing, and I was inspired by the way they welcomed risk taking and outside-the-box thinking.

During a chat over coffee, one of my new officemates told me he’d owned a successful business but lost everything overnight when his company went bankrupt. Afterward, he



“I found the fast-moving, entrepreneurial environment exhilarating.”

picked himself up and started over. Now, he was the director of another successful company selling a completely different product. His positive attitude about the experience made me realize it’s OK if you hit a dead end—you just shift gears and move in a new direction.

So I took a gamble, deciding that I would use my sabbatical to change research direction. I figured I had nothing to lose. Instead of writing grants to support my new interest—a strategy that hadn’t worked in the past—I decided to write a conceptual paper, hoping to gain credibility in the new field.

The topic I zeroed in on—glacial events that covered all of Earth—had captured my imagination when I was an undergraduate student. But I hadn’t read much about it since then, so I spent a lot of time

catching up on the literature. Then, I formulated an idea about how mass glacial events might have transformed the oceans, and I outlined it in a paper.

Against all odds, the paper was accepted by a top disciplinary journal. After it was published, I received something I wasn’t accustomed to: an invitation to present the idea at a major international conference. There, I met the scientist who pioneered the theory I’d written about, which stimulated collaborations that catapulted me into new areas of science. That was 5 years ago, and I’ve been much more productive and satisfied with my job ever since.

As academic researchers, we’re often funneled into increasingly narrow areas of specialization as we progress in our careers—but that doesn’t mean we’re trapped. Business moguls reinvent their careers all the time. Academics can, too. ■

Thomas Gernon is an associate professor at the University of Southampton in the United Kingdom.