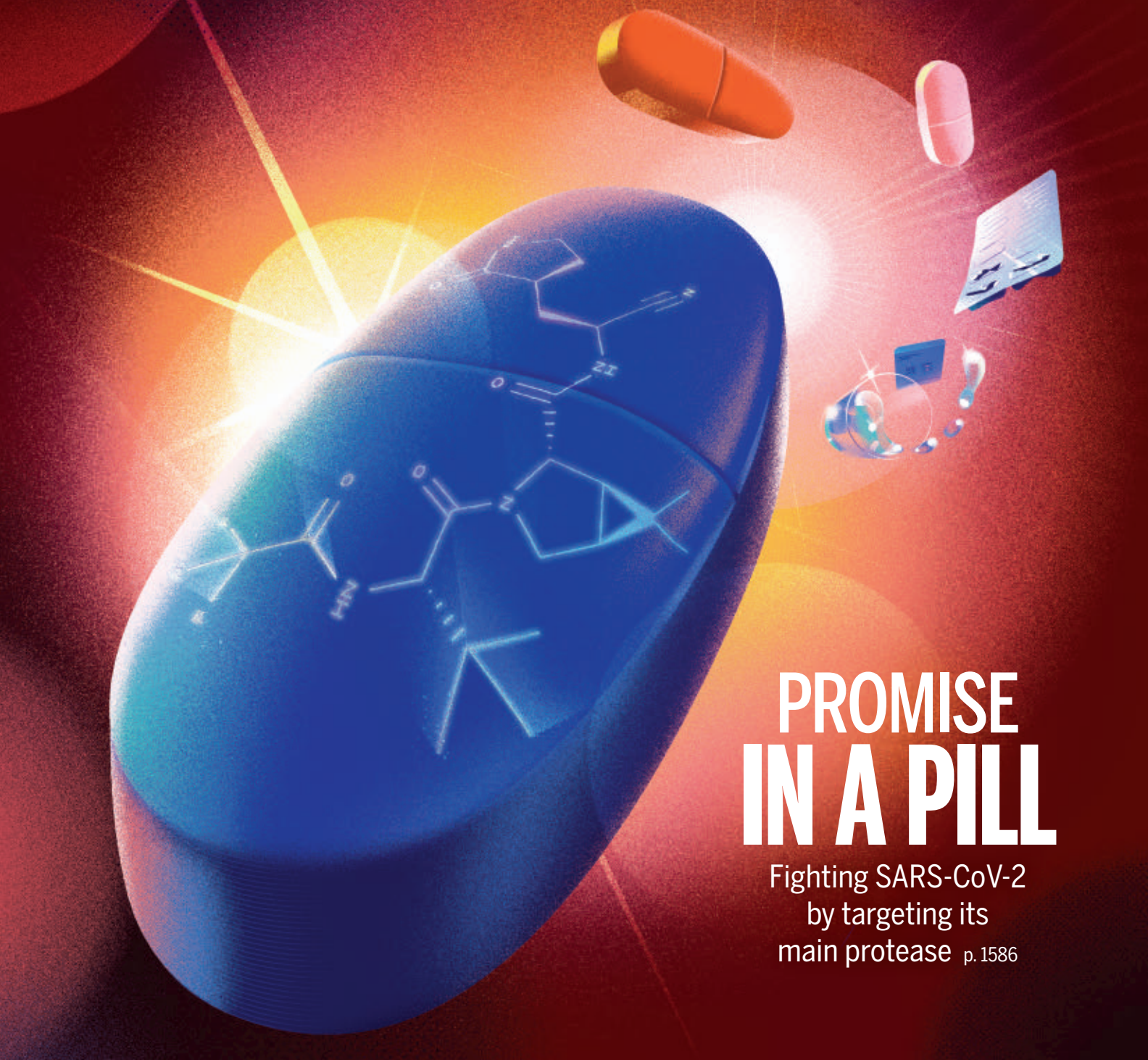
What to watch for in science during 2022 p. 1540

Giant ichthyosaur from the Triassic pp. 1554 & 1578

Tilt-stabilized hybrid perovskites p. 1598

Science

$15
24 DECEMBER 2021
science.org
AAAS



PROMISE IN A PILL

Fighting SARS-CoV-2 by targeting its main protease p. 1586



Apply for our exciting research Prize!

SCIENCE-PINS PRIZE FOR NEUROMODULATION

SCIENCE-PINS

$25, 000 Grand Prize!
Get published in *Science*!

The *Science*-PINS Prize is a highly competitive international prize that honors scientists for their excellent contributions to neuromodulation research. For purposes of the Prize, neuromodulation is any form of alteration of nerve activity through the delivery of physical (electrical, magnetic, or optical) stimulation to targeted sites of the nervous system with impications for translational medicine.

For full details, judging criteria and eligibility requirements, visit:

www.science.org/pins

Submissions Open: 15 December 2021

Science AAAS

PINS 品驰

Science Translational Medicine AAAS



Workshop speakers. Top row (L to R): Dinshaw Patel, Bing Zhu ,Yang Shi. Bottom row (L to R): Karolin Luger, Cynthia Wolberger, Peter Jones.

Tsinghua-*Science* Workshops: Epigenetics Structure and Function

In a recent series of online workshops organized by Tsinghua University and Science/*AAAS, scientists explored alterations in gene activity caused by mechanisms beyond changes to the genetic sequence.*

Modifying DNA with organic additions, such as a methyl group, can change the function of the molecule. Such epigenetic changes impact a wider range of biological processes, from DNA maintenance and immune-cell memory to gene regulation and disease. Scientists use a variety of methods–including structural biology, gene knockouts, and more–to explore the impact of epigenetic processes.

On September 23, 2021, Haitao Li–associate director of the Beijing Advanced Innovation Center for Structural Biology and associate dean of the School of Medicine, Tsinghua University–opened the session on structural and functional work on DNA methylation and its regulation.

The first speaker, Dinshaw Patel, a structural biologist at Memorial Sloan Kettering Cancer Center in New York, described the work of various investigators on DNA methyltransferases. As he explained, conformational changes to these molecules modulate their performance of various functions, such as DNA maintenance, which can impact human disease.

The next speaker, Bing Zhu, investigator and deputy director of the Institute of Biophysics at the Chinese Academy of Sciences, discussed his laboratory's work on DNA methylation. For example, he explained how changes in DNA methylation patterns arise in aging and tumorigenesis. He also described the role of methylation in creating memory B cells in the immune system, observing, "We have profound transcriptional memory in memory B cell activation."

On October 21, 2021, Yang Shi–who discovered the first histone demethylase, LSD1 (lysine-specific demethylase 1)–explained how this work precipitated potential cancer treatments. Now a principal investigator at the Ludwig Institute for Cancer Research at the University of Oxford, United Kingdom, he studies the roles of chromatin and RNA modifications in cancer. He remarked, "We're very interested in understanding immune checkpoint blockage through manipulating chromatin regulators." For example, he showed that LSD1 can make a tumor more responsive to immune therapy. In addition, he has used RNA interference to study the function of genes in eukaryotes.

Continuing the discussion about chromatin, Karolin Luger–the Jennie Smoly Caruthers Endowed Chair of Biochemistry at the University of Colorado Boulder–talked about the evolution of nucleosomes, which "are targets of epigenetic control," she said. Finding histone-like proteins in the single-celled archaea allowed her team to compare eukaryotic and noneukaryotic chromatin. Using gene knockout studies, her lab showed that these histone-like proteins are crucial for infectivity by a virus.

In the last session, held on December 2, 2021, Cynthia Wolberger–professor and director of the Department of Biophysics and Biophysical Chemistry at the Johns Hopkins University School of Medicine–described the role of reversible posttranslational modifications of histone residues in gene regulation. For example, she showed how large modifications, such as adding a ubiquitin molecule to a histone, impact transcription regulation. She delved deeper into the structural and biochemical evidence for how ubiquitination of histone H2B stimulates specific methyltransferases, and concluded with a discussion of the role of impaired regulation of H2B-specific deubiquitinating enzyme inhibitors in cancer.

To close this lecture series, cancer researcher Peter Jones, chief scientific officer at the Van Andel Institute in Grand Rapids, Michigan, updated participants on the methylation of DNA in nucleosomes. These processes depend on a specific combination of methyltransferases and accessory proteins. Employing cryogenic electron microscopy and gene knockouts, Jones has depicted the structure and function of such complexes. He concluded, "It's been a very interesting project in which we relate human disease to biological processes to chromatin studies and nucleosome function."

As these lectures demonstrate, epigenetics can be harnessed to study the basic mechanisms of biology as well as elucidate disease processes. Such studies could even be used to develop treatments for various diseases, including cancers.

Sponsored by

FRCBS 北京生物结构前沿研究中心 Beijing Frontier Research Center for Biological Structure

IMAGES PROVIDED BY TSINGHUA UNIVERSITY

CONTENTS

24 DECEMBER 2021 • VOLUME 374
ISSUE 6575



1554 & 1578

NEWS

IN BRIEF

1540 Areas to watch: What's coming up in 2022

IN DEPTH

1543 Early lab studies shed light on Omicron's behavior
Some data hint the new variant may be milder—but most scientists reserve judgment
By G. Vogel and K. Kupferschmidt

1544 Omicron sparks a vaccine strategy debate
Current boosters, variant-specific shots, and universal coronavirus vaccines all have fans *By J. Cohen*

1546 Will global vaccine shortfalls end in 2022?
Novavax vaccine authorization bolsters COVAX's optimistic new supply forecast *By J. Cohen*

1547 NIH's Collins steps down: 'I won't miss the nasty politics'
Geneticist's 12-year tenure marked by funding increases, new initiatives, and battling the COVID-19 pandemic
By J. Kaiser

1548 With double pulsar, theory of gravity passes toughest test yet
Study 16 years in the making shows how gravity distorts orbits and slows down photons
By D. Clery

1549 Early migration may have spread Celtic languages
Ancient DNA reveals 3000-year-old influx of people from France to Britain
By A. Gibbons

FEATURES

1550 Amazon delivery
A rainforest expedition brings COVID-19 vaccines to some of Brazil's most remote and vulnerable groups
By S. Moutinho

INSIGHTS

PERSPECTIVES

1554 Early and fast rise of Mesozoic ocean giants
A whale-sized ichthyosaur shows how fast these reptiles evolved
By L. L. Delsett and N. D. Pyenson
RESEARCH ARTICLE p. 1578

1555 Magnetic control in the terahertz
Strong magnon-phonon coupling may help develop superfast optical drives
By D. M. Juraschek and P. Narang
REPORT p. 1608

1557 A new tool to probe SARS-CoV-2 variants
Virus-like particles offer a new way to investigate genetic variation in SARS-CoV-2
By B. A. Johnson and V. D. Menachery
REPORT p. 1626

1558 Teaching natural enzymes new radical tricks
Cytochromes P450 were engineered to conduct abiotic, stereoselective radical reactions
By Y. Zhang and P. Dydio
REPORT p. 1612

1559 Fibrin sparks inflammation in the oral mucosa
Fibrin deposits in the oral mucosa trigger neutrophil activation and bone destruction *By T. Vicanolo and A. Hidalgo*
RESEARCH ARTICLE p. 1575

1561 COVID-19 vaccine breakthrough infections
Vaccine efficacy wanes over time but can be fully restored with a booster dose
By R. K. Gupta and E. J. Topol

POLICY FORUM

1563 Enabling commercial success of industrial biotechnology
Commercial-scale research translation has been muted *By B. W. Biggs* et al.

BOOKS *ET AL.*

1566 Revisiting paleontology's greatest hits
An otherwise solid dinosaur primer does little to convey the changing face of the field *By R. Black*

1567 Dispatches from the front lines
A new film documents the fraught first days of the COVID-19 pandemic in New York City
By R. S. Krauss

ILLUSTRATION: STEPHANIE ABRAMOWICZ/DINOSAUR INSTITUTE, NATURAL HISTORY MUSEUM OF LOS ANGELES COUNTY

LETTERS

1568 Governments must halt vulture poisoning
By P. I. Plaza and S. A. Lambertucci

1568 Transgender rights rely on inclusive language
By M. Miyagi et al.

1569 Brazil's deception threatens climate goals
By L. Ferrante and P. M. Fearnside

RESEARCH

IN BRIEF

1572 From *Science* and other journals

RESEARCH ARTICLES

1575 Immunology
Fibrin is a critical regulator of neutrophil effector function at the oral mucosal barrier
L. M. Silva et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABL5450
PERSPECTIVE p. 1559

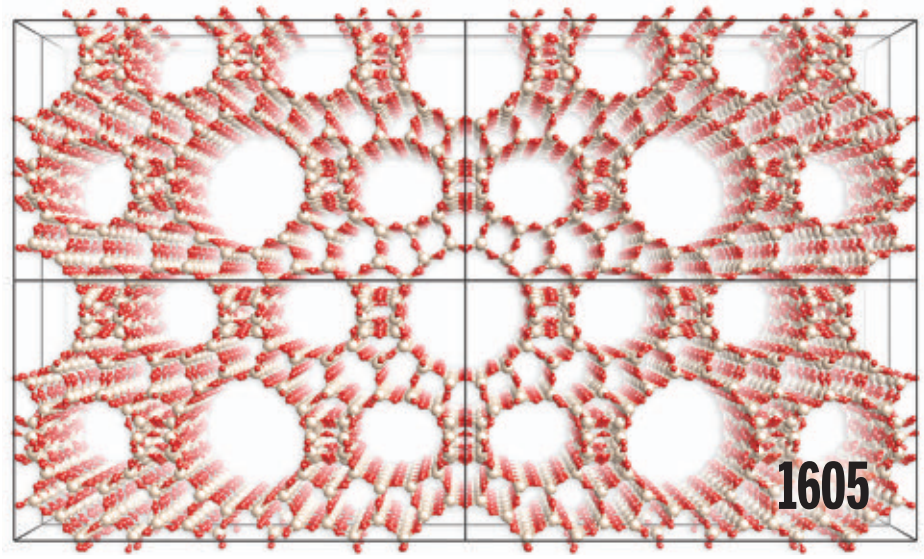
1576 Human evolution
The energetics of uniquely human subsistence strategies
T. S. Kraft et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABF0130

1577 Plant science
Cell-by-cell dissection of phloem development links a maturation gradient to cell specialization *P. Roszak* et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABA5531

1578 Paleontology
Early giant reveals faster evolution of large body size in ichthyosaurs than in cetaceans
P. M. Sander et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABF5787
PERSPECTIVE p. 1554

1579 Molecular biology
Pol IV and RDR2: A two-RNA-polymerase machine that produces double-stranded RNA
K. Huang et al.

1586 Coronavirus
An oral SARS-CoV-2 M^{pro} inhibitor clinical candidate for the treatment of COVID-19
D. R. Owen et al.



1593 Electrochemistry
Enhancement of lithium-mediated ammonia synthesis by addition of oxygen
K. Li et al.

REPORTS

1598 Perovskite structure
Stabilized tilted-octahedra halide perovskites inhibit local formation of performance-limiting phases
T. A. S. Doherty et al.

1605 Zeolites
A stable aluminosilicate zeolite with intersecting three-dimensional extra-large pores *Q.-F. Lin* et al.

1608 Ultrafast magnetism
Terahertz light–driven coupling of antiferromagnetic spins to lattice
E. A. Mashkovich et al.
PERSPECTIVE p. 1555

1612 Biocatalysis
Stereodivergent atom-transfer radical cyclization by engineered cytochromes P450 *Q. Zhou* et al.
PERSPECTIVE p. 1558

1616 Carbon nanotubes
Semiconductor nanochannels in metallic carbon nanotubes by thermomechanical chirality alteration
D.-M. Tang et al.

1621 Coronavirus
Molecular basis of immune evasion by the Delta and Kappa SARS-CoV-2 variants *M. McCallum* et al.

1626 Coronavirus
Rapid assessment of SARS-CoV-2–evolved variants using virus-like particles *A. M. Syed* et al.
PERSPECTIVE p. 1557

1632 Immunotherapy
Dietary fiber and probiotics influence the gut microbiome and melanoma immunotherapy response *C. N. Spencer* et al.

DEPARTMENTS

1537 Editorial
It's not as easy as it looks *By H. Holden Thorp*

1642 Working Life
A lesson in failure *By Jennifer Lanni*

ON THE COVER

Artist's rendering of the path from a positive COVID-19 test to a potential drug treatment. Researchers have developed a drug to target SARS-CoV-2, the virus that causes COVID-19. The molecule, known as PF-07321332, blocks the virus's main protease, an enzyme critical to its replication. The molecule can be delivered in pill form, which may lead to a COVID-19 treatment that can be used outside of a hospital setting. See page 1586. *Illustration: Jason Solo/The Jacky Winter Group*



CREDIT: Q.-F. LIN *ET AL.*

Science Staff 1536
AAAS News & Notes 1570
Science Careers 1641

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 © 2021 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: Air assist delivery: $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.science.org; bulk rate on request. **Authorization to reproduce** material for internal or personal use under circumstances not falling within the fair use provisions of the Copyright Act can be obtained through the Copyright Clearance Center (CCC), www.copyright.com. The identification code for *Science* is 0036-8075. *Science* is indexed in the *Reader's Guide to Periodical Literature* and in several specialized indexes.

The HKIAS Rising Star Lecture Series nurtures leading young researchers

5 fascinating lectures on mathematics, chemistry, physics, materials science, and life sciences

*The **Hong Kong Institute for Advanced Study (HKIAS)** recognizes the promise of 15 scientists who are just beginning their careers. "Nurturing new talent with a rewarding career has been a major mission of the Hong Kong Institute for Advanced Study since our establishment in 2015," said Jacob C. Huang, executive director of HKIAS. "We want to celebrate the accomplishments and promise of researchers at the early stages of their independent careers." To do that, the institute launched the HKIAS Rising Star Lecture Series in 2021.*

In the second half of 2021, 15 promising scholars at the City University of Hong Kong (CityU) presented their latest work in five **HKIAS Rising Star Lectures**, focusing on mathematics, chemistry, physics, materials science, and life sciences (See "HKIAS Rising Stars and Their Research Presentations"). The goal of these lectures—sponsored in part by the Kwang Hua Educational Foundation—is to promote diversity in academia and to draw international attention to the latest work from the next generation of academic leaders. "This is a great promotional opportunity for CityU and our Rising Star scholars," Huang said. HKIAS offered the top rising stars grants that could be used to conduct their research projects at CityU.

Mathematics

Kicking off the HKIAS Rising Star Lecture Series on September 15, three young mathematicians explored new approaches to navigating complexities with models. **Junhui Wang**—a specialist in statistical machine learning—started the series with a discussion of latent factor modeling, which can be used to improve the analytical predictability and accuracy of relational data. Statistical mechanics expert **Pierre Nolin** discussed how self-organized criticality can explain the complexity in many natural phenomena, including modeling forest fires. To complete this session, **Xianpeng Hu**—who studies partial differential equations—explored advances in understanding compressible versions of the well-known Navier–Stokes equations that describe the motion of viscous fluids.

Chemistry

The September 30 lecture saw three experts introduce their latest findings in interface chemistry, biological chemistry, and intercalation chemistry. Materials scientist **Angus Hin Lap Yip** explored the many crucial applications of metal halide perovskites, which make up a recent class of semiconductors that can be tuned to improve the optoelectrical properties of solar cells and LEDs. Next, **Chun Kit Kwok**—cofounder of the Hong Kong RNA Club—described the fascinating world of RNA structures and highlighted how mapping and targeting of functional RNA G-quadruplexes and their interactions with proteins might be applied to treat diseases such as cancers. **Chaoliang Tan** closed this lecture with an update on structural engineering of layered nanomaterials by lithiation intercalation chemistry, which can be applied to photothermal cancer therapy and aqueous zinc-ion batteries.

HKIAS Rising Stars and their research presentations

Mathematics



Junhui Wang

Professor, Department of Mathematics and School of Data Science

"Latent Factor Model: Methodology, Theory and Applications"

Pierre Nolin

Associate Professor, Department of Mathematics

"Forest Fires and Self-Organized Criticality"





Xianpeng Hu

Associate Professor, Department of Mathematics

"Existence and Concentration of Weak Solutions for Compressible Navier–Stokes Equations"

Chemistry

Angus Hin Lap Yip

Professor, Department of Materials Science and Engineering and School of Energy and Environment

"Molecularly Engineered Interfaces in Metal Halide Perovskite Optoelectronic Materials and Devices"





Chun Kit Kwok

Associate Professor, Department of Chemistry and State Key Laboratory of Marine Pollution

"A Renaissance in RNA Structure"

Chaoliang Tan

Assistant Professor, Department of Electrical Engineering

"Structural Engineering of Layered Nanomaterials by Lithium Intercalation Chemistry"



IMAGES: HKIAS

HKIAS Rising Stars and their research presentations

Physics



Danfeng Li
Assistant Professor, Department of Physics
"Soft-Chemistry Approaches to Superconductivity in Infinite-Layer Nickelates"

Qi Liu
Assistant Professor, Department of Physics
"High-Voltage Lithium Cobalt Oxide Cathode Materials for Future Smart Phones"





Cheng Wang
Assistant Professor, Department of Electrical Engineering and Member of the State Key Laboratory of Terahertz and Millimeter Waves
"Scaling Up Lithium Niobate Photonics for Future Optoelectronics"

Materials science

Yang Lu
Professor, Department of Mechanical Engineering and Department of Materials Science and Engineering
"Diamond Semiconductor: Deep Elastic Strain Engineering"





Dangyuan Lei
Associate Professor, Department of Materials Science and Engineering
"Plasmonic Nanocavity for Light Manipulation and Energy Conversion"

Tao Yang
Assistant Professor, Department of Materials Science and Engineering
"Chemically Complex Intermetallic Alloys: A New Frontier for Innovative Materials"



Life sciences



Gigi Pui Chi Lo
Associate Professor, Department of Biomedical Sciences
"Development of Nanophotosensitizing Systems for Targeted Photodynamic Therapy"

Liang Zhang
Assistant Professor, Department of Biomedical Sciences
"Dissecting Protein Interactome in Health and Diseases"





Kwok On Lai
Assistant Professor, Department of Neuroscience
"RNA Transport in Neuron and Why It Matters"

IMAGES: HKIAS

Physics

Condensed matter physics, in particular the properties of superconducting materials, is an area of intensive study at CityU. To open the Physics series on October 19, **Danfeng Li** described his role in discovering the world's first infinite-layer nickelate superconductors through a soft-chemistry approach and how these materials might allow zero electrical resistance, which could revolutionize high-performance electric power transmission. For more on power, battery expert **Qi Liu** explained how changes in using lithium cobalt oxide could improve a battery's energy capacity but not destroy its structure. Wrapping up this lecture, microelectronics engineering specialist **Cheng Wang** showcased his success in enhancing the light–matter interactions in nanophotonic structures, such as integrated lithium niobate photonic circuits, which can be applied to optical communications and nonlinear optics.

Materials science

Materials science, the study of the structure and properties of solid materials, is a key research field for CityU and HKIAS. Opening the November 5 lecture, **Yang Lu**—an expert on the nanomechanics of the hardest natural material, diamond—highlighted diamond's potential as a future semiconductor and how microfabrication and deep-strain engineering can optimize diamond's electronic properties for use in microelectronics, quantum-information technologies, and optoelectronics. In describing the confinement of light on the nanoscale in optical cavities, **Dangyuan Lei** used plasmonic and photonic low-dimensional materials and structures as examples and revealed their unprecedented potential applications, such as using dielectric microcavities to construct a waterproof perovskite microlaser and symmetry-broken particle-on-film nanocavities to boost the second-harmonic conversion efficiency of noble metals. Last in this session, **Tao Yang** revealed his passion for developing new materials for advanced engineering applications by introducing the audience to a family of chemically complex intermetallic alloys.

Life sciences

In the final lecture of the HKIAS Rising Star Series on November 24, three researchers explained how basic research in the life sciences might be translated to patients. As an example, **Gigi Pui Chi Lo**—a specialist in photosensitizing agents—talked about advances in nanophotosensitizers that could improve photodynamic therapy in cancer patients. Another expert in cancer biology and proteomics, **Liang Zhang**, proposed the CRISPR-assisted RNA–protein interaction detection method to delineate RNA–protein interactions in live cells, which could lead to the development of new biomarkers and therapeutics. Closing out the series, neuroscientist **Kwok On Lai** updated attendees on how the motor protein kinesin mediates intracellular transport to control the formation of neuronal connections. This work could answer long-standing questions around dendritic messenger RNA transport and translation, two important processes disrupted in various brain disorders.

Sponsored by

CityU 香港城市大學 City University of Hong Kong
Hong Kong Institute for Advanced Study

Phone: +852-3442-6611
Email: hkias@cityu.edu.hk

Science AAAS

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. Hurtley (UK), Phillip D. Szuromi, Sacha Vignieri **SR. EDITORIAL FELLOW** Andrew M. Sugden (UK) **SR. EDITORS** Gemma Alderton (UK), Caroline Ash (UK), Brent Grocholski, Pamela J. Hines, Di Jiang, Priscilla N. Kelly, 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, Yuen Yiu, Laura M. Zahn **ASSOCIATE EDITORS** Michael A. Funk, Bianca Lopez, Seth Thomas Scanlon (UK), Yury V. Suleymanov **LETTERS EDITOR** Jennifer Sills **LEAD CONTENT PRODUCTION EDITORS** Harry Jach, Lauren Kmec **CONTENT PRODUCTION EDITORS** Amelia Beyna, Jeffrey E. Cook, Chris Filiatreau, Julia Haber-Katris, Nida Masiulis, Abigail Shashikanth, 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), Alice Whaley (UK), Anita Wynn **PUBLICATIONS ASSISTANTS** Alexander Kief, Ronmel Navas, Isabel Schnaidt, Brian White **EXECUTIVE ASSISTANT** Jessica Slater **ASI DIRECTOR, OPERATIONS** Janet Clements (UK) **ASI SR. OFFICE ADMINISTRATOR** Jessica Waldock (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 Clery (UK), Jon Cohen, Jeffrey Mervis, Elizabeth Pennisi **ASSOCIATE EDITORS** Jeffrey Brainard, Kelly Servick, Catherine Matacic **NEWS REPORTERS** Adrian Cho, Jennifer Couzin-Frankel, Jocelyn Kaiser, Rodrigo Pérez Ortega (Mexico City), Robert F. Service, Erik Stokstad, Paul Voosen, Meredith Wadman **INTERN** Rachel Fritts **CONTRIBUTING CORRESPONDENTS** Warren Cornwall, Andrew Curry (Berlin), Ann Gibbons, Sam Kean, Eli Kintisch, Kai Kupferschmidt (Berlin), Andrew Lawler, Mitch Leslie, Eliot Marshall, Virginia Morell, Dennis Normile (Tokyo), Elisabeth Pain (Careers), Charles Piller, Gabriel Popkin, Michael Price, Joshua Sokol, Richard Stone, Emily Underwood, Gretchen Vogel (Berlin), Lizzie Wade (Mexico City) **CAREERS** Rachel Bernstein (Editor), Katie Langin (Associate Editor) **COPY EDITORS** Julia Cole (Senior Copy Editor), Morgan Everett, Cyra Master (Copy Chief) **ADMINISTRATIVE SUPPORT** Meagan Weiland

Creative Director Beth Rakouskas

DESIGN MANAGING EDITOR Marcy Atarod **INTERIM GRAPHICS MANAGING EDITOR** Chris Bickel **PHOTOGRAPHY MANAGING EDITOR** William Douthitt **WEB STRATEGY MANAGER** Kara Estelle-Powers **MULTIMEDIA MANAGING PRODUCER** Joel Goldberg **DESIGN EDITOR** Chrystal Smith **DESIGNER** Christina Aycock **GRAPHICS EDITOR** Nirja Desai **INTERACTIVE GRAPHICS EDITOR** Kelly Franklin **SENIOR GRAPHICS SPECIALISTS** Holly Bishop, Nathalie Cary **SENIOR SCIENTIFIC ILLUSTRATOR** Valerie Altounian **SCIENTIFIC ILLUSTRATOR** Ashley Mastin **SENIOR PHOTO EDITOR** Emily Petersen **PHOTO EDITOR** Kaitlyn Dolan **SOCIAL MEDIA STRATEGIST** Jessica Hubbard **SOCIAL MEDIA PRODUCER** Sabrina Jenkins **WEB DESIGNER** Jennie Pajerowski **SENIOR PODCAST PRODUCER** Sarah Crespi **VIDEO PRODUCER** Meagan Cantwell

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 **MANAGER, BUSINESS ANALYSIS** Cory Lipman **BUSINESS ANALYST** Kurt Ennis **FINANCIAL ANALYST** Isacco Fusi **ADVERTISING SYSTEM ADMINISTRATOR** Tina Burks **DIGITAL/PRINT STRATEGY MANAGER** Jason Hillman **SENIOR MANAGER, PUBLISHING AND CONTENT SYSTEMS** Marcus Spiegler **ASSISTANT MANAGER DIGITAL/PRINT** Rebecca Doshi **SENIOR CONTENT AND PUBLISHING SYSTEMS SPECIALIST** Jacob Hedrick **SENIOR CONTENT SPECIALISTS** Steve Forrester, 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 Dhere

ASSOCIATE DIRECTOR, BUSINESS DEVELOPMENT Justin Sawyers **GLOBAL MARKETING MANAGER** Allison Pritchard **DIGITAL MARKETING MANAGER** Aimee Aponte **JOURNALS MARKETING MANAGER** Shawana Arnold **MARKETING ASSOCIATES** Ashley Hylton, Mike Romano, Tori Velasquez, Jenna Voris, Justin Wood **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 MANAGMENT** Kris Bishop **PRODUCT DEVELOPMENT MANAGER** Scott Chernoff **PUBLISHING TECHNOLOGY MANAGER** Michael Di Natale **SR. PRODUCT ASSOCIATE** Robert Koepke **PRODUCT ASSOCIATE** Anne Mason **SPJ ASSOCIATE MANAGER** Samantha Bruno Fuller **SPJ ASSOCIATE** Casey Buchta

MARKETING MANAGER Kess Knight **BUSINESS DEVELOPMENT MANAGER** Rasmus Andersen **SENIOR INSTITUTIONAL LICENSING MANAGER** Ryan Rexroth **INSTITUTIONAL LICENSING MANAGER** Marco Castellan, Claudia Paulsen-Young **SENIOR MANAGER, INSTITUTIONAL LICENSING OPERATIONS** Judy Lillibridge **SENIOR OPERATIONS ANALYST** Lana Guz **FULFILLMENT COORDINATOR** Melody Stringer

DIRECTOR, GLOBAL SALES Tracy Holmes **US EAST COAST AND MID WEST SALES** Stephanie O'Connor **US MID WEST, MID ATLANTIC AND SOUTH EAST SALES** Chris Hoag **US WEST COAST SALES** Lynne Stickrod **ASSOCIATE DIRECTOR, ROW** Roger Goncalves **SALES REP, ROW** Sarah Lelarge **SALES ADMIN ASSISTANT, ROW** Victoria Glasbey **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** Yoshimi Toda (Tokyo), Miyuki Tani (Osaka)

DIRECTOR, COPYRIGHT, LICENSING AND SPECIAL PROJECTS Emilie David **RIGHTS AND LICENSING COORDINATOR** Jessica Adams **RIGHTS AND PERMISSIONS ASSOCIATE** Elizabeth Sandler **LICENSING ASSOCIATE** Virginia Warren

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
science.org/authors/
science-information-authors

REPRINTS AND PERMISSIONS
science.org/help/
reprints-and-permissions

MEDIA CONTACTS
scipak@aaas.org

MULTIMEDIA CONTACTS
SciencePodcast@aaas.org
ScienceVideo@aaas.org

INSTITUTIONAL SALES AND SITE LICENSES
science.org/librarian

PRODUCT ADVERTISING & CUSTOM PUBLISHING
advertising.science.org/
products-services
science_advertising@aaas.org

CLASSIFIED ADVERTISING
advertising.science.org/
science-careers
advertise@sciencecareers.org

JOB POSTING CUSTOMER SERVICE
employers.sciencecareers.org
support@sciencecareers.org

MEMBERSHIP AND INDIVIDUAL SUBSCRIPTIONS
science.org/subscriptions

MEMBER BENEFITS
aaas.org/membership/benefits

AAAS BOARD OF DIRECTORS
CHAIR Claire M. Fraser
PRESIDENT Susan G. Amara
PRESIDENT-ELECT Gilda A. Barabino
TREASURER Carolyn N. Ainslie
CHIEF EXECUTIVE OFFICER Sudip Parikh
BOARD Cynthia M. Beall, Rosina M. Bierbaum, Ann Bostrom, Janine Austin Clayton, Laura H. Greene, Kaye Husbands Fealing, Maria M. Klawe, Robert B. Millard, William D. Provine

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

BOARD OF REVIEWING EDITORS (Statistics board members indicated with **(S)**)

Erin Adams, *U. of Chicago*
Takuzo Aida, *U. of Tokyo*
Leslie Aiello, *Wenner-Gren Fdn.*
Deji Akinwande, *UT Austin*
Judith Allen, *U. of Manchester*
Marcella Alsan, *Harvard U.*
Sebastian Amigorena, *Inst. Curie*
James Analytis, *UC Berkeley*
Trevor Archer, *NIEHS, NIH*
Paola Arlotta, *Harvard U.*
David Awschalom, *U. of Chicago*
Delia Baldassarri, *NYU*
Nenad Ban, *ETH Zürich*
Nandita Basu, *U. of Waterloo*
Franz Bauer, *Pontificia U. Católica de Chile*
Ray H. Baughman, *UT Dallas*
Carlo Beenakker, *Leiden U.*
Yasmine Belkaid, *NIAID, NIH*
Philip Benfey, *Duke U.*
Kiros T. Berhane, *Columbia U.*
Joseph J. Berry, *NREL*
Alessandra Biffi, *Harvard Med.*
Chris Bowler, *École Normale Supérieure*
Ian Boyd, *U. of St. Andrews*
Emily Brodsky, *UC Santa Cruz*
Ron Brookmeyer, *UCLA* **(S)**
Christian Büchel, *UKE Hamburg*
Dennis Burton, *Scripps Res.*
Carter Tribley Butts, *UC Irvine*
György Buzsáki, *NYU School of Med.*
Mariana Byndloss, *Vanderbilt U. Med. Ctr.*
Annmarie Carlton, *UC Irvine*
Simon Cauchemez, *Inst. Pasteur*
Ling-Ling Chen, *SIBCB, CAS*
M. Keith Chen, *UCLA*
Zhijian Chen, *UT Southwestern Med. Ctr.*
Ib Chorkendorff, *Denmark TU*
Amander Clark, *UCLA*
James J. Collins, *MIT*
Robert Cook-Deegan, *Arizona State U.*
Virginia Cornish *Columbia U.*
Carolyn Coyne, *Duke U.*
Roberta Croce, *VU Amsterdam*
Ismaila Dabo, *Penn State U.*
Jeff L. Dangl, *UNC*
Chiara Daraio, *Caltech*
Nicolas Dauphas, *U. of Chicago*
Christian Davenport, *U. of Michigan*
Frans de Waal, *Emory U.*
Claude Desplan, *NYU*
Sandra Díaz, *U. Nacional de Córdoba*
Ulrike Diebold *TU Wien*
Stefanie Dimmeler, *Goethe-U. Frankfurt*
Hong Ding, *Inst. of Physics, CAS*
Dennis Discher, *UPenn*
Jennifer A. Doudna, *UC Berkeley*
Ruth Drdla-Schutting, *Med. U. Vienna*
Raissa M. D'Souza, *UC Davis*
Bruce Dunn, *UCLA*
William Dunphy, *Caltech*
Scott Edwards, *Harvard U.*
Todd Ehlers, *U. of Tübingen*
Andrea Encalada, *U. San Francisco de Quito*
Nader Engheta, *UPenn*
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.*
Jay Gallagher, *U. of Wisconsin*
Daniel Geschwind, *UCLA*
Ramon Gonzalez, *U. of South Florida*
Sandra González-Bailón, *UPenn*
Nicolas Gruber, *ETH Zürich*
Hua Guo, *U. of New Mexico*
Taekjip Ha, *Johns Hopkins U.*
Sharon Hammes-Schiffer, *Yale U.*
Wolf-Dietrich Hardt, *ETH Zürich*
Louise Harra, *U. Coll. London*
Carl-Philipp Heisenberg, *IST Austria*
Ykä Helariutta, *U. of Cambridge*
Janet G. Hering, *Eawag*
Christoph Hess, *U. of Basel & U. of Cambridge*
Heather Hickman, *NIAID, NIH*
Hans Hilgenkamp, *U. of Twente*
Janneke Hille Ris Lambers, *ETH Zürich*
Kai-Uwe Hinrichs, *U. of Bremen*
Deirdre Hollingsworth, *U. of Oxford*
Randall Hulet, *Rice U.*
Auke Ijspeert, *EPFL*
Akiko Iwasaki, *Yale U.*
Stephen Jackson, *USGS & U. of Arizona*
Erich Jarvis, *Rockefeller U.*
Peter Jonas, *IST Austria*
Matt Kaeberlein, *U. of Wash.*
William Kaelin Jr., *Dana-Farber Cancer Inst.*
Daniel Kammen, *UC Berkeley*
Kisuk Kang, *Seoul Nat. U.*
Sabine Kastner, *Princeton U.*
V. Narry Kim, *Seoul Nat. U.*
Robert Kingston, *Harvard Med.*
Nancy Knowlton, *Smithsonian Institution*
Etienne Koechlin, *École Normale Supérieure*
Alex L. Kolodkin, *Johns Hopkins U.*
Julija Krupic, *U. of Cambridge*
Paul Kubes, *U. of Calgary*
Gabriel Lander, *Scripps Res.* **(S)**
Mitchell A. Lazar, *UPenn*
Wendell Lim, *UCSF*
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.*
David Lyons, *U. of Edinburgh*
Fabienne Mackay, *QIMR Berghofer*
Anne Magurran, *U. of St. Andrews*
Asifa Majid, *U. of York*
Oscar Marín, *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.*
Heidi McBride, *McGill U.*
C. Robertson McClung, *Dartmouth*
Rodrigo Medellín, *U. Nacional Autónoma de México*
C. Jessica Metcalf, *Princeton U.*
Baoxia Mi, *UC Berkeley*
Tom Misteli, *NCI, NIH*
Alison Motsinger-Reif, *NIEHS, NIH* **(S)**
Suresh Naidu, *Columbia U.*
Danielle Navarro, *U. of New South Wales*
Daniel Nettle, *Newcastle U.*
Daniel Neumark, *UC Berkeley*
Beatriz Noheda, *U. of Groningen*
Helga Nowotny, *Vienna Sci. & Tech. Fund*
Rachel O'Reilly, *U. of Birmingham*
Pilar Ossorio, *U. of Wisconsin*
Andrew Oswald, *U. of Warwick*
Isabella Pagano, *Istituto Nazionale di Astrofisica*
Elizabeth Levy Paluck, *Princeton U.*
Jane Parker, *Max Planck Inst. Cologne*
Giovanni Parmigiani, *Dana-Farber Cancer Inst.* **(S)**
Daniel Pauly, *U. of British Columbia*
Ana Pêgo, *U. do Porto*
Samuel Pfaff, *Salk Inst.*
Julie Pfeiffer, *UT Southwestern Med. Ctr.*
Philip Phillips, *UIUC*
Matthieu Piel, *Inst. 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.*
Trevor Robbins, *U. of Cambridge*
Joeri Rogelj, *Imperial Coll. London*
Amy Rosenzweig, *Northwestern U.*
Mike Ryan, *UT Austin*
Miquel Salmeron, *Lawrence Berkeley Nat. Lab*
Nitin Samarth, *Penn State U.*
Erica Ollmann Saphire, *La Jolla Inst.*
Joachim Saur, *U. zu Köln*
Alexander Schier, *Harvard U.*
Wolfram Schlenker, *Columbia U.*
Susannah Scott, *UC Santa Barbara*
Anuj Shah, *U. of Chicago*
Vladimir Shalaev, *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 Sivilotti, *U. Coll. London*
Alison Smith, *John Innes Ctr.*
Richard Smith, *UNC* **(S)**
John Speakman, *U. of Aberdeen*
Tara Spires-Jones, *U. of Edinburgh*
Allan C. Spradling, *Carnegie Institution for Sci.*
V. S. Subrahmanian, *Northwestern U.*
Ira Tabas, *Columbia U.*
Eriko Takano, *U. of Manchester*
Patrick Tan, *Duke-NUS Med. School*
Sarah Teichmann, *Wellcome Sanger Inst.*
Rocio Titiunik, *Princeton U.*
Shubha Tole, *Tata Inst. of Fundamental Res.*
Maria-Elena Torres Padilla, *Helmholtz Zentrum München*
Kimani Toussaint, *Brown U.*
Barbara Treutlein, *ETH Zürich*
Jason Tylianakis, *U. of Canterbury*
Wim van der Putten, *Netherlands Inst. of Ecology*
Ivo Vankelecom, *KU Leuven*
Henrique Veiga-Fernandes, *Champalimaud Fdn.*
Reinhilde Veugelers, *KU Leuven*
Bert Vogelstein, *Johns Hopkins U.*
Julia Von Blume, *Yale School of Med.*
David Wallach, *Weizmann Inst.*
Jane-Ling Wang, *UC Davis* **(S)**
Jessica Ware, *Amer. Mus. of Natural Hist.*
David Waxman, *Fudan U.*
Chris Wikle, *U. of Missouri* **(S)**
Terrie Williams, *UC Santa Cruz*
Ian A. Wilson, *Scripps Res.* **(S)**
Hao Wu, *Harvard U.*
Wei Xie, *Tsinghua U.*
Yu Xie, *Princeton U.*
Jan Zaanen, *Leiden U.*
Kenneth Zaret, *UPenn School of Med.*
Bing Zhu, *Inst. of Biophysics, CAS*
Xiaowei Zhuang, *Harvard U.*
Maria Zuber, *MIT*

EDITORIAL

It's not as easy as it looks



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

Amid the many miscommunications and misunderstandings about how to respond to the COVID-19 pandemic, scientists have been called on to do a better job of explaining their work to the public. In the United States, John Holdren, former science adviser to President Obama, has gone so far as to call for every scientist to become a trained communicator in an army of ambassadors for the whole country. That all sounds good on the surface. But just what do we expect these scientific emissaries to accomplish? Is it education? Is it advocacy? Is it changing behaviors?

Media training for scientists has been fashionable for years. Private companies cash in on the trend. Universities regularly conduct science communication workshops for researchers. There are conferences on the science of science communication. Science journalists have written books about how to convey complex information to a general audience. What more can be done?

It's not just a matter of translating jargon into plain language. As Kathleen Hall Jamieson at the University of Pennsylvania stated in a recent article, the key is getting the public to realize that science is a work in progress, an honorably self-correcting endeavor carried out in good faith. Moreover, scientists need to have some understanding of their audience to improve the chance of a true dialog. They may need to learn to listen and "read the room," and prepare different approaches for different audiences. Some communication lessons were learned, for example, by the US Centers for Disease Control and Prevention's 2014 Tips campaign for smoking cessation. Rather than focusing on the data showing why smoking is dangerous, the campaign revolves around personal stories of patients who have suffered smoking-related illnesses that individuals can relate to. Application of this principle—which meets folks where they are rather than burying them with data—led to remarkable progress in smoking cessation. Understanding this is a high-level skill that requires expertise.

It is also the case that not every scientist wants to take time away from research to be a voice for science. Traits that make a good researcher, such as concentration on details and laser focus on a problem, don't often carry over to the public stage. Most scientists prefer to persuade by performing meticulous, credible work.

I work with some of the best science communicators in the world, and I see how hard they have endeavored to hone their craft. This is a profession and a full-time job—not something that can be picked up in a workshop. Recently, I talked to one scientist who seems to succeed in wearing both hats. Rebecca Schwarzlose, an accomplished neuroscientist at the Washington University School of Medicine, wrote an outstanding popular book, *Brainscapes*, about neural maps. Her training as both a scientist and communicator didn't come easy. After getting her PhD from the Massachusetts Institute of Technology, Schwarzlose initially left the lab and became an editor at *Trends in Cognitive Sciences*. After hearing her give a talk to a general audience, I complimented her on the clarity of her message and asked her if she thought every scientist could do what she has done. "No, I don't," she said. "I think what we should probably be aiming for is finding ways to give scientists tools to help them communicate, not necessarily even with a general audience, but just more broadly between different disciplines." That sounds good, but it's not the same thing as preparing a cadre of communicators to go to Kiwanis clubs full of climate deniers.

> **"It's not just a matter of translating jargon into plain language."**

After talking to Schwarzlose, I think we need to be careful with the notion that every scientist can be easily trained to also be an outstanding communicator. You frequently hear the idea that a science communication course should be a required part of graduate training. This can't hurt, but asking someone to be a skilled science communicator after taking one course is like asking someone who has taken a course in chemistry to discover a novel reaction. Truly well-trained science communicators—individuals who devote their lives to helping the public understand research—deserve more respect from their research colleagues. As Schwarzlose's story shows, it's not something that just anyone can pick up quickly. Maybe a better idea is to figure out how to improve the partnership between researchers and these public communicators.

–H. Holden Thorp

PHOTO: CAMERON DAVIDSON

10.1126/science.abn7633



A new paradigm in AI research and education

While artificial intelligence (AI) can be a powerful technology for addressing complex grand challenges such as climate change, public health, food and water security, and economic prosperity, there is still no cohesive way to organize all the methodologies used by AI in generating practical solutions and to properly evaluate the science (and scientists) of AI. Bringing academic rigor, reproducibility, and standardization to the field is, in part, the mission of a professor who heads up the world's first AI-focused graduate university.

In 2019, the government of the United Arab Emirates (UAE) made a historic decision to launch Mohamed Bin Zayed University of Artificial Intelligence (MBZUAI). Based in Abu Dhabi, the university has a simple but bold mission: establish and continually evolve interdisciplinary, collaborative research and development capability in the field of AI, while educating innovators and leaders with the breadth and depth of knowledge to grow technology and enterprise in the UAE and globally.

To achieve this ambitious, holistic directive, the board of directors selected Dr. Eric Xing, an international leader in AI research and entrepreneurship, to serve as its inaugural president.

Prior to his appointment, Xing was a professor of computer science at Carnegie Mellon University in Pittsburgh, Pennsylvania, and the founder, chairman, and chief scientist of Petuum Inc., a pioneering company that builds standardized AI development platforms and operating systems for industrial AI applications.

Xing pioneered the development of fundamental technologies in machine learning, including metric learning, graphical models, and time-evolving and sample-specific networks. More recently, he propelled the emergence of the new field of SysML—systems and machine learning—and the technology for controlled generation of texts.

"Inside the academic community, everyone is trying to do something exciting with AI, but there are limits in the current environment, particularly the lack of agility and flexibility in industrial collaboration, tech transfer, and entrepreneurship," he says.

"There is also tension between applied and fundamental research, both in terms of peer recognition and financial reward. Coming to MBZUAI gave me the opportunity to start from scratch and think about big questions, like finding the optimal system for management and administrative support to create a healthy, vibrant organization able to do the most creative and effective science. I am inspired to experiment with new ideas and open opportunities in research, education, and entrepreneurship," Xing adds.

But he has concerns about the speed and "the somewhat chaotic development" of the AI field.

In Xing's conception, "many things need to be rethought. This goes beyond a choice of scientific approach and into style, taste, and public influence." He encourages consideration of what a university can offer to enable the best minds to do the best work, and of what type of AI research the academic community, rooted in the university, is in a unique position to advance.

"Only by clarifying the scientific principles underlying AI, and solving the problems of reproducibility and standardization of AI production, can the real 'engineering era of AI' start. This is the same path trodden by many highly mature disciplines."

Xing believes that a clearer theoretical framework will give innovation a better foundation, ensuring that AI implementations have a more

IMAGE COURTESY OF MBZUAI

generalizable and universal engineering framework, and offering new research a more purposeful direction. "A research institution must aspire to provide the perfect intellectual environment for these types of explorations, and this ambition and responsibility is at the core of the academic mission of this university," he explains.

At MBZUAI, institutional research is focused around four themes representing significant touchpoints of AI in society: (1) services and quality of life, with an emphasis on improving services offered by the government and industry; (2) industrial and manufacturing technologies, addressing how AI improves and catalyzes productivity, manufacturing, and industrial processes; (3) future game changers, bolstering emerging sectors and entrepreneurial ventures; and (4) sustainability of vital resources and the environment, examining how AI can optimize and thereby conserve the environment.



Dr. Eric Xing

Currently, there are three departments at MBZUAI: machine learning, computer vision, and natural language processing, with plans to expand. There are intentionally no silos. "Faculty are encouraged and empowered to be collaborative," says Xing. "We are building an organization that is not very hierarchical. It's like the ancient School of Athens where good ideas are exchanged, shared, and nurtured."

Xing's priorities include recruiting top international faculty and developing educational pipelines. The work is a combination of investments in three areas: (1) fundamental theories and algorithms such as unified learning frameworks, and logical and causal inferences; (2) foundational platforms and programs, in particular new backend systems for automatically composing, training, and tuning large-scale machine learning models and pipelines; and (3) applied systems, especially those relevant to the local economy and regional needs in clean energy and health care.

He also is creating a sustainable, growth-oriented entrepreneurial and economic ecosystem to support the advancement of scientific and technical capabilities and the placement of talent. Xing is ensuring that the MBZUAI tech transfer team communicates with industry leaders and faculty to perform appropriate needs assessments.

"With the traditional academic system, it is challenging to be agile enough to accommodate these types of opportunities and initiatives," he says. "MBZUAI is reprogramming more effective mechanisms to optimize creativity and sustainable growth and to allow faculty to ask ambitious and deep questions."

Naturally, education takes precedence. MBZUAI has Ph.D. and Master's degrees as well as masterclasses and customized education programs for executives. Future expansion will add professional training programs for engineers, Bachelor's degrees, and other continuing education certificates, all with a unique and actionability-driven infusion of AI general and domain knowledge.

"We have the opportunity to build a program that is relevant to the next stage of people's careers, with the right balance of knowledge and industry experience," says Xing. A mandated internship offers a taste of what it's like to work in industry, and students can choose whether to produce a thesis or capstone project based on an AI application.

But AI science is only the first step. MBZUAI's goal is to bolster the UAE economy through strategic economic development, workforce development, and even policy contributions, all influenced by and built upon a platform of AI investments.

"We want to build economic diversity, and AI presents this opportunity," says Xing. "AI is still very young. The UAE leadership see this as a great opportunity to enter the world stage with a timely and wise investment."

Recently, MBZUAI was commissioned to provide a "crash course" on AI technology, policy, ethics, and risk for government representatives and leaders. "We are not only providing a research and development hub for AI in the academic sense," adds Xing, "We are also providing a think tank for leadership and a cradle of training functionality for the local talent."

When he was being recruited, Xing asked the question: Why was it important for the government of Abu Dhabi to establish a university of AI here? The answer, he says with a chuckle, significantly contributed to his decision to come to MBZUAI.

"The leaders of the UAE are among the most forward looking and visionary worldwide. They look beyond their current economic needs and situation to energy and economic needs 50 years in the future. The leadership and citizens here have the sense of urgency to stay ahead of the curve, to be prepared and be competitive. They know AI is the driving force of the next industrial revolution."

Sponsored by

MOHAMED BIN ZAYED
UNIVERSITY OF
ARTIFICIAL INTELLIGENCE

IMAGE COURTESY OF MBZUAI

NEWS

IN BRIEF

Edited by **Jeffrey Brainard**



More than 660,000 white flags were placed in a temporary art installation on the National Mall in Washington, D.C., in September to commemorate Americans who died of COVID-19.

PHOTO: PATRICK SEMANSKY/AP PHOTO

AREAS TO WATCH: WHAT'S COMING UP IN 2022

COVID-19 looms large over the next year in science

Any look at the immediate future of science must start with the pandemic coronavirus. As the world enters the third year of COVID-19, researchers are striving for a handle on what the latest variant of concern, Omicron, may mean for the pandemic's trajectory (see p. 1543). Given the sheer number of infections globally, further variants are likely to arise in 2022. Because a large part of the global population now has some level of immunity from vaccines or infection, scientists are bracing for variants much better at evading human immune responses. Whether vaccines must be adapted is still unclear (see p. 1544). At the same time, scientists are developing a new generation of vaccines that may give broader immunity or elicit a stronger response in the mucosae of the respiratory tract. In 2022, the world could also begin to see the impact of oral antiviral drugs targeting SARS-CoV-2, which could reduce stress on health systems by reducing deaths and the number of people who need hospitalization. Researchers will look to crack the mysteries of Long Covid, in which people experience debilitating symptoms weeks or months after their infection clears. And as global vaccine supplies increase, a key question will be whether poor countries can play catchup and get large swaths of their populations immunized (see p. 1546). In the rest of this section, *Science*'s news staff forecasts other areas of research and policy likely to make headlines in the coming year.

U.S. redefines research security

POLICY | Congress and the White House could soon finalize legislation and regulations designed to balance openness and national security in the funding of academic research. With U.S. policymakers worried that federal research dollars could fuel technological advances in China, Congress appears ready to ban federally funded scientists from participating in its foreign talent recruitment programs and to tighten oversight of any type of research support from that economic giant. Several U.S. scientists have already been accused of failing to disclose their ties to China, and the outcome of their trials could determine the fate of the government's China Initiative, a 3-year-old law enforcement effort that critics say has unfairly targeted scientists of Chinese descent and criminalized bookkeeping errors.

Ion beam to conjure rare nuclei

NUCLEAR PHYSICS | The fleeting atomic nuclei normally forged only in stellar explosions will find a home on Earth after the $730 million Facility for Rare Isotope Beams (FRIB) fires up at Michigan State University. The world's most powerful ion source, the linear accelerator can fire any nucleus—from hydrogen's single proton to uranium atoms' massive core—into targets to produce new, unstable nuclei. It aims to yield 80% of all theoretically possible isotopes, including more than 1000 that have never been observed. With the FRIB, physicists hope to bolster their understanding of the structure of nuclei, decipher how stellar explosions produce heavy elements, and search for new forces of nature.

China debuts exascale computers

INFORMATION TECHNOLOGY | In 2022, China is expected to demonstrate the world's two fastest and most powerful computers, which recent reports indicate have surpassed a long-sought performance milestone. At a supercomputing conference in late November, news emerged that China has built the first "exascale" supercomputers, able to carry out more than 1 quintillion (10^{18}) calculations per second. China has yet to officially announce the machines, for reasons that remain unclear. And details about their performance have yet to appear on the TOP500 list of supercomputers, which ranks the world's top machines based on common benchmarks. But according to technology watchers, the new supercomputing champs are Sunway Computer Co.'s OceanLight and the National Supercomputing Center's Tianhe-3. Installation of what will be the first U.S. exascale computer, Frontier, is underway at Oak Ridge National Laboratory; it is slated to come online in 2022. Exascale supercomputers are expected to enable the marriage of artificial intelligence with massive data sets, transforming fields such as personalized medicine and materials discovery and generating more realistic models of climate change and the accelerating expansion of the universe.

Bevy of landers heads for Moon

SPACE SCIENCE | Fifty years after humans last set foot on the Moon, robotic missions are returning en masse—and reviving dreams of human exploration. Following successful Chinese rover landings, three NASA-funded robotic landers, developed by small startups, will launch in 2022, and they could be joined by probes from Russia, Japan, and India. NASA's program has two purposes: to conduct research, especially on the extent and availability of the Moon's trace water, and to pave the way for human exploration by cheaply delivering payloads to the dusty surface. The coming year will also mark the first orbital launches for two mammoth rockets capable of taking astronauts and heavy payloads to the Moon and beyond: NASA's overdue and costly Space Launch System and SpaceX's Starship.



NASA's Orion spacecraft and Space Launch System rocket, standing 98 meters tall, face prelaunch tests.

PHOTO: NASA

U.N. panel to tackle pollution?

ENVIRONMENT | The United Nations Environment Assembly is set to vote in February 2022 on a proposal to create a scientific advisory body to study the risks of chemical pollution and waste, modeled on the Intergovernmental Panel on Climate Change. The United Nations already has several conventions on certain types of pollution, such as mercury and persistent organic chemicals. But proponents of the new panel say there's a need for broad assessments to help policymakers spot emerging problems and identify research needs. More than 1800 scientists have signed a petition of support to be presented at the assembly.

Listening for spiraling black holes

ASTROPHYSICS | Gravitational wave detectors have registered the collisions of star-size black holes, but scientists are also mounting a different kind of search for bigger prey. In 2022, they hope to have enough data to clearly discern the low-frequency rumble of gravitational waves from pairs of supermassive black holes, with masses billions of times that of the Sun, as they spiral toward merger. To detect the spiraling pairs, observers train large radio telescopes on dozens of pulsars—collapsed stars emitting radio beams that, as the pulsar spins, appear as pulses with clocklike regularity. Tiny variations in the beats would signal passing gravitational waves that stretch the space between the pulsars and Earth. Although these pulsar timing arrays cannot pinpoint individual black hole binaries, teams in Europe and North America say their arrays have detected a faint signal that may be the background rumbling of spiraling pairs across the universe.

U.S. research innovators launch

FUNDING | The two biggest U.S. research agencies are poised to create units that would implement congressional mandates to accelerate practical discoveries. The National Institutes of Health could get up to $3 billion in 2022 to stand up an Advanced Research Projects Agency for Health that would fund "breakthrough" approaches to treating a host of diseases.

The National Science Foundation could get up to half that amount for a new technology directorate that would focus on speeding up the commercial application of academic research funded by NSF's ongoing programs. However, those increases could shrink—or even disappear—if lawmakers extend a freeze on all federal agency budgets, now in place until at least mid-February 2022, for the rest of the fiscal year.

Biodiversity pact to get stronger

CONSERVATION | Safeguards for endangered species could get a boost if nations adopt a new framework for the Convention on Biological Diversity, to be in effect until 2050. Under a plan developed by negotiators for a meeting in 2022 in China, the 196 nations involved would seek to expand protection of natural ecosystems, emphasize sustainability, ensure that profits from using genetic resources are shared equitably, and raise at least $700 million by 2030 to fund these efforts. Targets include conserving 30% of the land and sea; reducing the spread of invasive species; halving the world's pollution, including reducing pesticide use by two-thirds and eliminating the discharge of plastic waste; and increasing city dwellers' access to "green and blue" spaces. The new plan aims to improve on a previous one, adopted in 2010, that set goals for 2020, which were not completely met. New approaches include establishing ways to monitor and report progress toward protecting species and ecosystems and doing more to involve stakeholders, such as Indigenous people, in making decisions.

China moves to allow GM crops

AGRICULTURE | China will likely approve the first commercial planting of genetically modified (GM) corn and soybeans, under proposed revisions to field trial and safety requirements unveiled in November. GM corn seeds could be on the market by the end of 2022. Currently, papaya is the only GM plant for human consumption approved for production within China. GM cotton is widely cultivated, and GM poplar is also available. China has funded research on GM corn and soybeans for more than 10 years, but public opposition and official caution kept the plants in the lab. The country imports large quantities of GM corn and soybeans for use in processed food and animal feed, which has led to calls from government officials to relax restrictions on growing GM varieties domestically. The government is unlikely to allow commercialization of GM rice soon because China is self-sufficient in the grain and consumers are wary about modifications to the country's most important staple food.

Methane hunter to orbit

GLOBAL WARMING | At November's climate summit in Glasgow, U.K., world leaders made a commitment to cut their emissions of methane, a powerful greenhouse gas, 30% by 2030. But to verify whether those cuts actually happen, a new generation of satellites targeting greenhouse gas emissions will begin to reach orbit in 2022. One is MethaneSAT, set for launch in October and developed by the Environmental Defense Fund, a nonprofit climate advocacy group. It is expected to sharpen the ability to detect methane plumes from sources such as rice paddies and leaking pipelines. Soon after, it will be joined by two satellites developed by the nonprofit Carbon Mapper project, targeting the world's worst emitters—not just of methane, but also of carbon dioxide, the primary gas driving global warming.

Malaria vaccine arrives in Africa

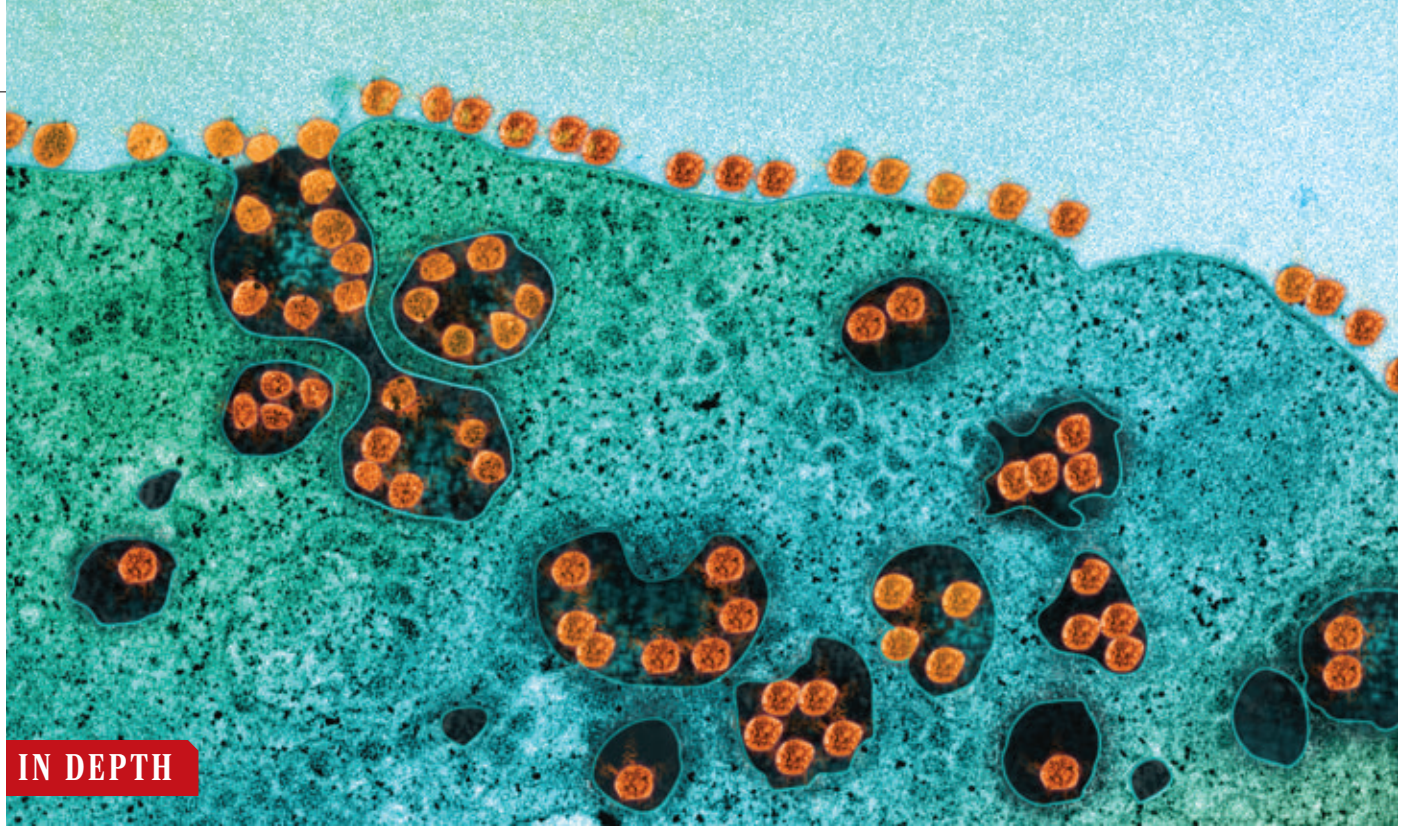
PUBLIC HEALTH | In 2022 for the first time, countries across Africa will be able to use a vaccine to protect children against malaria, which still kills more than 260,000 of those under the age of 5 there each year. The RTS,S vaccine, in development for 3 decades, finally received approval from the World Health Organization in October, and in December, Gavi, the Vaccine Alliance, said it would spend more than $155 million buying the vaccine and supporting its rollout through 2025. The vaccine offers imperfect protection, lowering rates of hospitalization for severe malaria by about 30%. But in initial trials, it reached many children who don't sleep under bed nets and are particularly vulnerable. Another study demonstrated that it is especially effective when combined with antimalaria medication given prophylactically during the high-risk rainy season. Researchers and public health advocates will carefully track where the vaccine is used and what impact it has on malaria morbidity and mortality. Scientists have yet to develop any vaccines against the different species of parasites that cause malaria outside sub-Saharan Africa.



A revised biodiversity agreement could target invasive species. The recent introduction of the California kingsnake (*Lampropeltis californiae*) to the Canary Islands reduced native reptiles by as much as 90%.

PHOTO: JORGE SAAVEDRA

SCIENCE.ORG/NEWS
Read more news from *Science* online.



IN DEPTH

Omicron virus particles (orange) amass in vesicles and are released on the surface of Vero cells. A study suggests Omicron replicates less well deep inside human lungs.

IMAGE: SCHOOL OF PUBLIC HEATH/DEPARTMENT OF PATHOLOGY/LKS FACULTY OF MEDICINE AND ELECTRON MICROSCOPE UNIT/UNIVERSITY OF HONG KONG

COVID-19

Early lab studies shed light on Omicron's behavior

Some data hint the new variant may be milder—but most scientists reserve judgment

By **Gretchen Vogel** *and* **Kai Kupferschmidt**

The new SARS-CoV-2 variant exploding around the planet is forcing humanity to adapt at a breakneck speed. Late last week, countries across Northern Europe imposed stringent new measures to try to bring down soaring case numbers. The Netherlands issued a near-complete lockdown of public life and allowed residents to receive no more than two guests per day—with an exception for Christmas and New Year's Eve, when four visitors are allowed. Denmark, which lifted all pandemic restrictions in September, closed theaters, concert halls, museums, and other public spaces and banned alcohol sales after 10 p.m.

Driving those decisions is alarm at Omicron's wildfirelike spread—and caution in the face of major uncertainty. If Omicron makes people just as sick as the Delta variant, models predict a staggering rise of hospitalizations—many times what most health systems can handle. If Omicron causes milder disease than Delta, things would be less catastrophic—but even then, "a considerable overload of the hospitals is to be expected," a group of experts warned in a 19 December report to the German government. A massive Omicron wave might also lead to many more cases of Long Covid.

It will take weeks before epidemiological data deliver a clear verdict about disease severity. But preliminary lab data, and a few real-world clues, are raising hope among optimists, while others reserve judgment. "If it turns out there is reduced severity from Omicron, there are now potentially some mechanisms that might help explain this," says Tom Peacock, a virologist at Imperial College London (ICL). "On the other hand, there may be some mechanisms that could explain the opposite. So it's still a bit of a mess until the real-world data crystallizes."

South Africa has seen only a modest increase in hospitalizations for severe COVID-19, even 1 month after cases started to explode there. "We're all still really cautious about that data, but it seems to be holding out," says virologist Wendy Burgers of the University of Cape Town. Moreover, early signs suggest South Africa's Omicron wave may not last long. Cases are already decreasing in Gauteng province, the outbreak's epicenter, which is a bit of a mystery given how transmissible Omicron seems to be, says Trevor Bedford, a bioinformatics specialist at the University of Washington, Seattle, and the Fred Hutchinson Cancer Research Center. He suspects part of the explanation is that more infections than usual went unnoticed because they were mild or asymptomatic, so the real peak was even larger than official statistics showed.

But early data from Europe, which has an older population than South Africa, are less hopeful. Based on early hospitalization data, for example, a report by modelers at ICL concluded there are "at most limited changes in severity compared with Delta." "I don't know how to square these two things, and that is probably giving me the most pause at the moment," Bedford says.

Lab data offer a little more hope. Antibodies resulting from vaccination or infection have been shown to lose most of their power against Omicron, but another immune system defense—T cells, which recognize and kill infected cells in the body—appears to do better. Burgers and her colleague Catherine Riou tested the response of T cells from people who had received one or two doses of the Johnson & Johnson COVID-19 vaccine or two doses of Pfizer. In all three groups, the T cell response to Omicron dropped by a modest amount compared with the original

strain of the virus—between 20% and 30%, the researchers estimate. That's much less than the drop in antibody efficacy. "T cells are holding out against Omicron," Burgers concludes. "And the data is very consistent across vaccines."

That may help explain South Africa's low hospital burden, Burgers says. "Of course it could be other things about the biology of this variant," she adds, "but certainly from everything we know about T cells, this is what they do—control a virus once you've been infected. So this is their time to shine."

MEANWHILE, RESEARCHERS studying Omicron's behavior in the lab are turning up surprises. "There is something very weird about this virus," Peacock says. Normally, SARS-CoV-2 replicates well in Vero cells—kidney cells from African green monkeys—genetically engineered to be studded with the virus' receptors, ACE2 and TMPRSS2. With any previous strain, such infections lead to visible areas of damage in the cell culture called plaques. But that's happening much less with Omicron, Peacock says. "It's not liking the cell lines that have been our bread and butter all the way through."

Researchers in Hong Kong have seen notable differences in another system, in which the virus infects healthy lung tissue taken from lung cancer patients during surgery. Omicron replicates much faster than previous variants in cells from the bronchi, but much slower in cells from lower in the lungs. "Whilst the other strains take off after about 72 hours in the bronchus, this one goes up within 24 hours," says John Nicholls, a pathologist at the University of Hong Kong. The data, published in a short note online and submitted to a journal, might explain why Omicron spreads better while perhaps causing less severe disease. Virus in the bronchi is more likely to be exhaled, whereas deep in the lung it is likely to cause more severe disease. But Nicholls urges caution in interpreting the data. "We're being very, very careful," he says.

Some of Omicron's many mutations initially didn't bode well. Scientists have zeroed in on changes close to the furin cleavage site, a motif in the spike protein that allows it to be more easily cleaved, which is an important step in the infection of a human cell. Mutations around this site appear to have made both Alpha and Delta more transmissible and better at infecting cells. Omicron carries the mutation Alpha had, P681H, and another one close by called N679K. "Just from sequence gazing, we thought, this is probably going to have Delta-like transmissibility and pathogenicity," Peacock says.

But a preprint by virologist Ravi Gupta of the University of Cambridge suggests the opposite could be true: Using viruses modified to carry the spike protein of Omicron and other variants, they found that Omicron's spike is cleaved much less efficiently than Alpha's or Delta's. That could explain why Omicron might be worse at infecting certain cells than previous variants.

His preprint also suggests another contrast. Research on previous variants has suggested SARS-CoV-2 can cause neighboring cells to merge into syncytia, large cells with multiple nuclei. Syncytia have been found in the lungs of patients who died of COVID-19 and may be related to disease severity, Gupta says. Kei Sato, a virologist at the University of Tokyo, has shown that the more pathogenic Delta variant was more likely to cause syncytia in cell culture. Now, Sato and Gupta both have data suggesting syncytia may be rarer with Omicron. Taking all of the data together, Gupta says, "I think the evidence is mounting … that the virus potentially causes less progression to severe disease."

The role of syncytia is far from clear, however. SARS-CoV, the virus that causes severe acute respiratory syndrome, has not been shown to cause them. And although SARS-CoV-2 seems to trigger syncytia in cell culture, they don't appear in animal models, Nicholls says: "In the hamsters we looked at, in the mice we looked at, you don't get the syncytia." Christian Drosten, a virologist at Charité University Hospital in Berlin, says syncytia in COVID-19 patients may have other explanations, such as the reactivation of herpes infections, which are known to cause syncytia formation.

PEACOCK SAYS it's simply too early to draw firm conclusions from the lab studies. "I think there's definitely something interesting going on there. But I wouldn't like to bet that other people are definitely going to find the exact same thing," he says. Gary Whittaker, a virologist at Cornell University, says Omicron seems to behave more like seasonal coronaviruses, which are often hard to grow in cell culture. That makes experiments on the new variant even harder to interpret, he says. "We've got to get a system which is much better matched to the virus, otherwise we will very easily get led astray if we're not careful."

Getting firm answers from the lab may well take until Easter, Drosten says. By that time, the world may have learned how much severe disease the variant causes the hard way. "So far, variants that replicate more have caused more disease," Drosten says. "So we would do well to be very careful with this variant." ■

COVID-19

Omicron sparks a vaccine strategy debate

Current boosters, variant-specific shots, and universal coronavirus vaccines all have fans

By **Jon Cohen**

Barely a year after vaccine developers reported spectacular successes against SARS-CoV-2, the latest variant of the virus has presented them with a new challenge. Omicron is spreading at unprecedented speed, and studies of blood samples and emerging real-world evidence show one reason is its ability to dodge immunity, whether from previous infections or vaccination. The new threat has galvanized vaccine developers and public health officials—but they are not all pushing in the same direction.

Reacting to evidence that an additional dose can restore some protection against Omicron, many countries are racing to give booster shots of existing vaccines. Some scientists think that is unnecessary, arguing two doses of the current messenger RNA (mRNA) vaccines may not prevent infections but will suffice to keep most people who are otherwise healthy from getting seriously sick with COVID-19. Others think it's time for revised vaccines. A booster tailored to Omicron would likely offer better protection, says vaccinologist Florian Krammer of the Icahn School of Medicine at Mount Sinai. Vaccine manufacturers, eyeing a new market, are doing the preliminary work to concoct new formulations.

Rather than chasing SARS-CoV-2 variants with tailored shots, another camp says it's better to push for universal vaccines that protect against all SARS-CoV-2 mutants, as well as other coronaviruses that nature throws at humans. Because of Omicron, "There is even stronger interest in our approach," says Kayvon Modjarrad of the Walter Reed Army Institute of Research, who is developing a pancoronavirus vaccine candidate that has just completed a phase 1 trial. Other pancoronavirus candidates may start trials in 2022.

The dramatic rise of the Omicron vari-

ant has brought a broader debate to a head: how to handle future COVID-19 waves as we move further down the Greek alphabet of SARS-CoV-2 variants. "The idea of reboosting every 6 months or even every year, that's not a public health strategy," says epidemiologist Céline Gounder of New York University's Grossman School of Medicine.

Paul Offit, a pediatrician at the Children's Hospital of Philadelphia, contends it's "extremely unlikely" Omicron will make healthy, vaccinated people under age 65 very sick because it can't outfox T cells—which early studies suggest are largely unaffected by this variant's mutations—and immune memory. The current vaccines, he notes, have shown 90% or better protection against severe disease, "as defined by the likelihood of having to visit a doctor, go to the hospital, or worse," he says.

In expecting SARS-CoV-2 vaccines to protect against transmission as well as mild and moderate disease, "we're holding this vaccine to a different standard," Offit says. He doesn't think giving booster shots to entire populations—as most countries that can afford it are now doing at breakneck speed—is necessary.

Gounder, though, argues a booster push currently makes sense because there are so many unknowns about Omicron's severity (see p. 1543). "Let's give third doses right now to buy us some time while we sort it out," Gounder says. It's also possible that boosting will reduce the incidence of Long Covid, a still-mysterious constellation of persistent symptoms that can arise after even a mild initial infection.

For now, vaccine boosters rely on the same formulations and doses authorized before Omicron arrived. But Moderna, which received authorization from the U.S. Food and Drug Administration (FDA) to cut its current dose in half for the third booster shot of its mRNA vaccine, is running a clinical trial of a full-dose booster to see whether it produces higher levels of antibodies against Omicron.

It and other vaccine developers are also creating variant specific updates of the currently authorized shots—which all are derived from the original SARS-CoV-2 that circulated in Wuhan, China, in early 2020. Moderna and the other producer of a widely used mRNA vaccine, Pfizer and its partner BioNTech, say they could begin to deliver an Omicron-tailored vaccine as soon as March. But they warn that shifting to a new vaccine would cut production of the current one.

Moderna has tested the impact of mRNA mixtures that combine the original variant's spike with new ones, and suspects this might provide broad enough protection to counter Omicron. Pfizer and BioNTech reported last week that in a clinical trial, a boost with an mRNA vaccine coding for the spike of an earlier variant, Alpha, elicited neutralizing antibody levels against Omicron more than four times higher than boosts with spikes from Beta, Delta, or an Alpha/Delta mix. "The Omicron variant shares multiple mutations with the Alpha variant," explained BioNTech's CEO, Uğur Şahin, at a press briefing.

Regulatory hurdles remain unclear: Will FDA or other regulatory bodies simply accept evidence of immune responses correlated with COVID-19 protection rather than wait perhaps months for clinical trial proof? An analysis of the Moderna's original vaccine efficacy trial found neutralizing antibodies accounted for only 68% of the protection. Still, Peter Marks, who heads the vaccine division at FDA, says, "Immunogenicity studies based on neutralization assays will suffice for a [vaccine] change if we need it."

Immunologist John Moore of Weill Cornell Medical College, who favors boosting as many people as possible with the current mRNA vaccines, calls it prudent to do the preliminary work on Omicron-focused vaccines, but suggests by the time they're scaled up, the Omicron wave may be over.

"Maybe we are too late for the first wave, but if Omicron is here to stay, we need a specific vaccine," Krammer counters. "I think not acting is problematic."

Moore also remains wary of companies, which could benefit financially from variant boosters, steering vaccine policy. "It is going to have to be based on as much data as possible, and some very serious discussions in the high levels of the administration, without the company execs putting their finger on the scales," he says.

Nicole Lurie, who heads the U.S. division of the nonprofit Coalition for Epidemic Preparedness Innovations (CEPI), also sees no need for an Omicron-specific shot given its rapid spread, but says the companies should be prepared to make vaccines for mutants that move more slowly. "Or they have to get faster at making vaccines," she says.

Lurie and others say Omicron has underscored the need for vaccines that protect against all severe acute respiratory syndrome (SARS)-related viruses—known collectively as sarbecoviruses—or, better yet, all viruses in the coronavirus family. CEPI in March put out a call for proposals to develop such vaccines, hoping to put $200 million toward such projects over the next 5 years. The U.S. National Institute of Allergy and Infectious Diseases (NIAID) in September announced grants totaling $36.3 million to university teams that have the same goal. "It's a compelling need," says NIAID Director Anthony Fauci, a co-author of a perspective last week on the topic in *The New England Journal of Medicine*.

A pancoronavirus vaccine may not be as tall an order as it might seem, says Fauci, who takes heart in the fact that current vaccines have powerfully prevented severe disease with all variants. "That bodes pretty well for the possibility of getting a pancoronavirus [vaccine] that you can boost with and protect against everything," he says. Already, several have fared well in animal trials.

Some vaccinated people may already have pan-sarbecovirus antibodies, a team led by Linfa Wang of the Duke-NUS Medical School in Singapore found. Wang's team has studied people who had survived SARS nearly 2 decades ago and recently received COVID-19 mRNA vaccines. In test tube studies, their antibodies neutralized the coronavirus that causes SARS, SARS-CoV-2, and several related viruses found in bats, the group reported in August.

Wang is now making a vaccine booster based on a hybrid of spike sequences from SARS-CoV and SARS-related bat viruses. But he doesn't have access to monkeys to rigorously test the concept first and still needs a corporate partner to back the project, he wrote in an email. "The Omicron emergence will hopefully make this move a bit faster!" ■



COVID-19 vaccine shots are prepared in Spain, one of many countries rushing to deliver booster doses to thwart the Omicron variant.

PHOTO: AP PHOTO/PAUL WHITE

COVID-19

Will global vaccine shortfalls end in 2022?

Novavax vaccine authorization bolsters COVAX's optimistic new supply forecast

By **Jon Cohen**

After facing setback after setback this year, the nonprofit formed to make sure COVID-19 vaccines reach the poorest countries of the world may finally live up to its promise in 2022. A 14 December report shows the COVID-19 Vaccines Global Access (COVAX) Facility is close to meeting a revised target of 1.42 billion available doses this year. And last week, the effort got a huge boost when the World Health Organization (WHO) gave its first emergency use listing (EUL) to a vaccine that COVAX is counting on for up to 1 billion doses next year.

Despite COVAX's recent successes and its optimistic new supply forecast, delivering the prized shots to needy countries isn't the last word in achieving global vaccine equity: Many nations may still struggle to distribute their supplies and, in some cases, overcome vaccine hesitancy. "Supply still needs attention, but we have pivoted to delivery and absorption as the main issues," says Seth Berkley, CEO of Gavi, the Vaccine Alliance.

Established by Gavi, WHO, UNICEF, and the Coalition for Epidemic Preparedness Innovations (CEPI), COVAX initially set a goal of vaccinating 20% of the population in every country—enough to cover health care workers and the people most at risk of developing severe disease. To accomplish this, COVAX said it needed to have 2 billion doses available by the end of this year. The latest COVAX forecasting report shows the effort will fall far short of that, with 1.38 billion doses available by year's end.

COVAX set out to buy vaccines in bulk and then provide them to 1.8 billion people in 92 low- and middle-income countries at little or no cost. But many wealthy countries ordered far more doses than they needed, bumping COVAX to the end of the purchasing line. COVAX also counted on the Serum Institute of India to supply up to 1.1 billion doses, but it backed out of the arrangement in March in order to protect India during its Delta variant surge.

Supply steadily increased as wealthy countries began to donate excess doses and more vaccine manufacturers received EULs from WHO, a COVAX requirement. Last month, Serum also began to export again. And on 17 December, WHO issued an EUL for the protein-based vaccine made by Serum on behalf of its U.S. biotech developer, Novavax.

The COVAX forecast projects it will have made 2.39 billion doses available by March 2022. It has options for a total of more than 6.5 billion doses by the end of next year. "I think we're at a tipping point of the problem being demand-driven versus supply-driven," says Nicole Lurie, U.S. director of CEPI.

For now, vaccine disparities between rich and poor remain stark. As of 16 December, 56.5% of the people in the world have received at least one of the 8.56 billion doses of COVID-19 vaccines administered, according to Our World in Data. But in low-income countries, that figure drops to 7.5%.

Half of COVAX's 92 priority countries now have access to enough doses to cover more than 20% of their population, Berkley notes. "Of course, it is not enough and there are still horrible inequalities."

The large numbers of doses soon to be "available" may not resolve inequities quickly. "The supply forecast number is certainly important to track, but it is not the most important measure of how many people are actually getting access to vaccines from COVAX," says Krishna Udayakumar, who heads the Duke Global Health Innovation Center.

Once manufacturers offer the vials, COVAX must allocate them, based on requests from countries and their ability to "absorb" the shipments. There's a 4- to 6-week lag between availability and delivery, which means at most, COVAX will deliver a total of 1 billion doses this year.

Delivery isn't the finish line, Udayakumar stresses. "Delivery for COVAX just means it showed up in an airport, and it takes several weeks if not months beyond that and an enormous amount of effort at the country level to get from airports to arms."

Udayakumar points to several, more nuanced issues that will likely continue to create problems for COVAX. High-income countries that donate vaccines often don't give recipient countries a timeline for delivery and sometimes provide vials that are near their expiration dates, making it difficult to administer doses in time, he says.

"The attempt to vaccinate the world is going to run into more intractable problems than the supply issue," agrees William Moss, an epidemiologist at Johns Hopkins Bloomberg School of Public Health.

Political instability afflicts countries that have the lowest COVID-19 vaccine coverage, whereas others have yet to express much demand. "There are many settings where COVID is just not the priority," Moss says.

Lurie says COVAX recipients have also become increasingly discriminating about the vaccines on offer even though WHO has given them all an authorization. "We're seeing countries say more and more 'We want the same vaccines used in America.'" Vaccine misinformation and the attendant skepticism and hesitancy have also become global challenges, she adds.

Berkley says COVAX is now moving away from trying to do a blanket distribution of vaccines to hit the 20% mark and instead is focusing on helping the 25 countries that have the lowest coverage. "We are creating bespoke plans for each one to deal with their specific bottlenecks," he says. ■



After a slow start, COVAX is increasing vaccine deliveries to nations such as Ivory Coast.

PHOTO: SIA KAMBOU/AFP VIA GETTY IMAGES



PHOTO: ANNA MONEYMAKER/THE NEW YORK TIMES/REDUX

U.S. SCIENCE POLICY

NIH's Collins steps down: 'I won't miss the nasty politics'

Geneticist's 12-year tenure marked by funding increases, new initiatives, and battling the COVID-19 pandemic

By **Jocelyn Kaiser**

Francis Collins ended his 12 years as director of the National Institutes of Health (NIH) on a high note last week, rushing to interviews and being showered with fond tributes from former presidents and even cellist Yo-Yo Ma. Collins, an affable, guitar-playing, motorcycle-riding geneticist and physician, steered the $43 billion agency through three administrations, won budget increases, and launched major new programs in cancer, neuroscience, and personalized medicine. He also led NIH's efforts to fight the coronavirus pandemic. Collins, who is returning to his lab at NIH's National Human Genome Research Institute, spoke with *Science* about some of his tenure's highs and lows. The interview has been edited for clarity and brevity.

Q: Diversity is one area where you wish you'd made more progress. But NIH recently ended a policy that some institutes hoped would help identify high-quality proposals from Black scientists that just missed the funding cutoff. Was that because of legal concerns?
A: Exactly. What we must not do in our efforts to try to right 400 years of wrongs is put forward solutions that can be readily attacked as inconsistent with Supreme Court rulings. We cannot run the risk of a legal challenge.

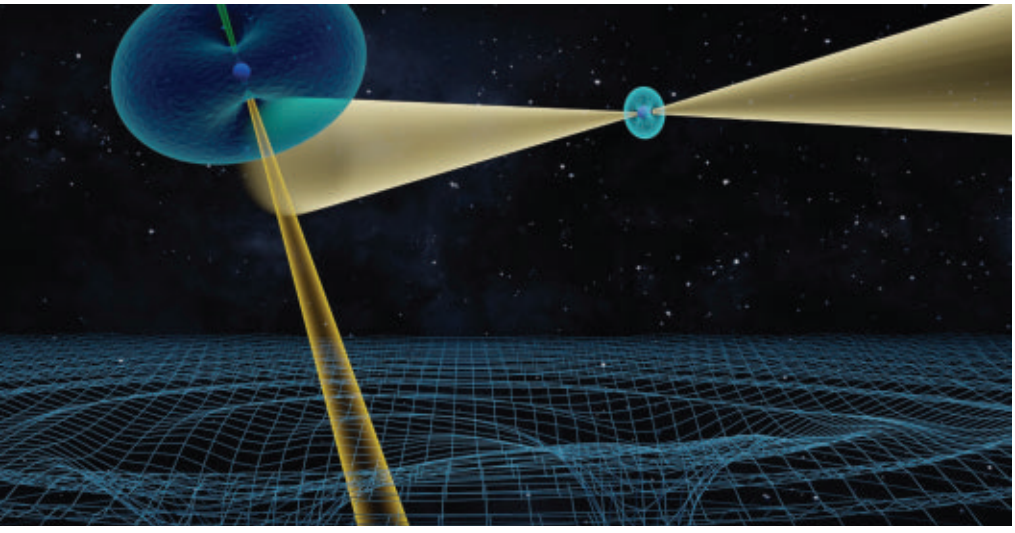
Q: How do you feel about your efforts to support younger, early-stage investigators?
A: We still have a problem that academic institutes keep trainees in graduate and postdoc positions too long. But that's not a lever that we can totally pull. In terms of the success rate for a first-time investigator, we have pushed that up to about 25%. Look at the numbers. We had 600 early-stage investigators funded in 2014. The number for [this year] is about 1400. I would say it's progress in the right direction.

Q: Do you think that ARPA-H, the new agency to accelerate biomedical research, is going to happen? And should it be a stand-alone agency rather than part of NIH?
A: I sure hope it happens. There seems to be enthusiasm in both the House [of Representatives] and the Senate. It's going to be quite a challenge to stand up a complicated organization like this. NIH is quite ready to do that while maintaining arm's length to ensure a very different culture at ARPA-H. It is going to have to figure out the most compelling projects that involve an opportunity to move something forward at a pace that otherwise wouldn't happen. It seems to me NIH is in a pretty good place to bubble up those kinds of projects.

Q: NIH has been involved in the Department of Justice's China Initiative, which aims to stop intellectual property theft, because of concerns that some NIH-funded scientists have not disclosed foreign funding. Some investigations have resulted in criminal prosecutions. Is this the right way to handle it?
A: We need to be sure when somebody is seeking substantial funding from NIH that they have disclosed their other support. It is clear there have been egregious incidents where people had substantial funding, even shadow labs in another country, [that they did not disclose]. It's up to the [federal prosecutors] to decide if somebody has broken a law, such as tax law.

Q: You came under fire from scientists in 2020 for suspending a grant to the EcoHealth Alliance and its subcontractor, the Wuhan Institute of Virology, to survey bat coronaviruses. NIH later restored it but with conditions that EcoHealth said were impossible to meet. Why haven't you relaxed those conditions and resumed the grant?
A: The termination of the grant, which came down as an order from the White House, was not something that we were given a choice about. [Since then,] it's clear that while the original justification for the work they were doing was quite compelling, there have been ways in which EcoHealth has not been responsive to NIH requests for information. For instance, they were supposed to notify NIH if a certain viral construct they were working with turned out to have a more than one log increase in infectivity for a humanized mouse, which did not happen. So they have not fully discharged their responsibilities working in an area that's obviously fraught with potential for human harm. [In imposing conditions on the grant,] we're treating them like we would any grantee that has responsibilities and hasn't lived up to them.

Q: What won't you miss about being NIH director?
A: I won't miss the nasty politics that unfortunately seem to surround a lot of what's happening right now as a reflection of the divisiveness in our country. I won't miss the slings and arrows being pitched at NIH, on the basis of political perspectives. I won't miss getting really horrible, hateful emails every time I show up on Fox News. And the things that they're quite comfortable saying about me and my family. I won't miss 100 hours a week [of work]. That oftentimes was utterly exhausting and consuming, but it needed to be. So I haven't had much chance to really stop, reflect, and think and maybe do some writing. I will be glad to be able to do some of those things. ■



The only known double pulsar system beams out radio waves that sweep past Earth metronomically.

IMAGE: © MICHAEL KRAMER/MPIFR

ASTROPHYSICS

With double pulsar, theory of gravity passes toughest test yet

Study 16 years in the making shows how gravity distorts orbits and slows down photons

By **Daniel Clery**

Since 2005 astronomers have monitored a pair of spinning neutron stars—two superdense objects whirling around each other every 2.5 hours while emitting metronomic radio pulses. That patient scrutiny has yielded the most wide-ranging tests yet of Albert Einstein's theory of gravity. Unsurprisingly, the general theory of relativity, as it's called, passed all the tests, leaving physicists both elated at the feat and disappointed that it yielded no clues to a "theory of everything" that would transcend general relativity.

The unusual astronomical system—the only known double pulsar—enabled the team to put general relativity to seven tests, including the most precise measurements yet of how ripples in spacetime called gravitational waves carry energy away from the stars, reducing their masses and causing their orbits to shrink. The system also displayed other effects that had never been measured before, including how photons from one of the stars slow down and bend as they pass through the intense gravitational field of the other.

"This is a really impressive suite of tests for general relativity—which of course passes," says Scott Ransom of the U.S. National Radio Astronomy Observatory, who was not involved in the study.

When the stellar duo was discovered in 2003, "We didn't realize what sort of a gem it was," says Michael Kramer of the Max Planck Institute for Radio Astronomy, who led the study. The stars, collapsed stellar remnants made of tightly packed neutrons, have masses greater than the Sun squeezed into balls the size of a city, generating intense gravitational fields. Their frenetic spins turn them into pulsars, emitting radio beams that sweep past Earth like a lighthouse—one of them every 2.8 seconds, the other every 23 milliseconds. Together, they act as two giant clocks whirling in each other's intense gravity. By looking for tiny variations in the pulses' timing, which reveal how gravity affects their orbits and the propagation of the photons themselves, researchers can put relativity to extreme tests.

To measure the pulses, Kramer and his colleagues used six large radio telescopes around the globe plus the Very Long Baseline Array, a set of 10 dishes spanning the United States. The fact that the team just kept on observing the pulsars month after month, for years, without publishing any results became a running joke among colleagues. It's been "a long time coming!" says Cherry Ng of the University of Toronto's Dunlap Institute for Astronomy & Astrophysics, who commends the team for its mammoth effort. "Combining data from multiple telescopes is not an easy task."

When the 56-page paper landed last week in *Physical Review X*, researchers felt it was worth the wait. "It's an incredible paper and is incredibly thorough," Ransom says. In total, the team says, general relativity agrees with the observations to a level of at least 99.99%.

The measurements were so precise that some couldn't be explained by general relativity's simpler predictions: for example, how gravity slows time, shifts light toward longer wavelengths, and ratchets the axis of elliptical orbits forward slightly on each revolution, an effect called precession. Theorists had to factor in other, more esoteric predictions of Einstein's theory, such as the way a dense, spinning mass twists the spacetime around it, and the way extreme gravity can distort the shape of an orbit. The latter is "a subtle, subtle effect that's never been measured before," says Ransom, who was impressed by just how many of these "higher order" effects they had to invoke.

Although confirming general relativity is satisfying, researchers want to find a chink in its armor one day. General relativity cannot be a complete description of how the universe works because, as far as anyone knows, it cannot be reconciled with that other pillar of modern physics, quantum mechanics. An unexplained gravitational aberration might hint at a path to a more complete theory. But the double pulsar results only make it harder to conjure up "other theories consistent with the data," says team member Thibault Damour of the Institute for Advanced Scientific Studies near Paris.

In addition to a testbed for relativity, the double pulsar may offer a window into the interiors of pulsars. To understand the behavior of the densely packed neutrons, researchers need to know a pulsar's density. Its mass can be determined easily from its orbit, but size—the other factor in density—is harder to gauge. A subtle coupling between the spin of each pulsar and its orbit, predicted by general relativity, could provide a way to calculate its radius—if the coupling could be measured accurately enough. Ransom, who calls pulsar size "a holy grail," says he had hoped the double pulsar researchers might be able to pin it down, but they were only able to set size limits. "It proved much trickier to pick out."

More patience will be needed. Ransom points out that in pulsar studies, the longer you observe, the better accuracy you achieve. In 10 or 20 more years, he says, the team may be able to bear down even further. "It gets much better with time." ■

ARCHAEOLOGY

Early migration may have spread Celtic languages

Ancient DNA reveals 3000-year-old influx of people from France to Britain

By **Ann Gibbons**

About 2900 years ago, an elderly woman was carefully buried with two lambs in her lap and a piece of chalk in her hand at a site now called Cliff's End Farm, about 30 kilometers north of Dover, U.K. She had been killed by sword blows to her skull, likely in a sacrificial rite. Nearby lay the bodies of a teenager, two children, and a man whose bones had been bundled along with a copper-tipped cow bone. Two of the dead had been born in Europe, according to the isotopes in their teeth. Now, a study adds new insight into their origins: They may have been part of a wave of early Celtic speakers to reach Britain.

Researchers report in *Nature* today that the genomes of these people and nearly 800 others document a previously unknown great migration from Europe that transformed the genetic makeup of people in southern Britain in the Late Bronze Age, 3100 to 2700 years ago. The migrants likely introduced Celtic languages related to those spoken today in Ireland, Wales, and Scotland. It was "the last major prehistoric migration to Britain, and it probably came from France," says archaeologist James Mallory of Queen's University Belfast. If so, Celtic languages began to reach Britain as much as 1000 years earlier than expected, he says.

For decades, prehistorians thought Celtic languages arrived in Britain about 2400 years ago in the Iron Age, along with "Celtic" art and inscriptions first spotted in Central Europe. But debate raged over the timing, the source of the languages, and whether they spread through migration or diffused on their own. "Without genetics, every theory was debated," Mallory says.

But now there are more ancient genomes from Britain than anywhere else. That is allowing researchers to trace the waves of people who reached the British Isles over the past 11,000 years. Hunter-gatherers came first, followed 6000 years ago by Early European Farmers (EEF) originally from Anatolia. About 4500 years ago, farmers with characteristic Bell Beaker pottery overran Britain, rapidly replacing 90% of inhabitants' DNA and swamping out most of the EEF ancestry.

In the new study—the largest ancient DNA study to date—a team of more than 200 researchers, led by population geneticist David Reich of Harvard University and archaeologist Ian Armit of the University of York, explored how the genetic makeup of British people evolved from about 6000 to 2000 years ago. The team analyzed about 1 million alleles across the genomes of 793 people who lived in Britain and Europe.



A woman (left) and man whose bones were bundled (top) were buried in a pit on the U.K. coast.

PHOTO: © WESSEX ARCHAEOLOGY

The data revealed an influx of people with more EEF markers than local Britons, starting about 4400 years ago. They began to mix gradually. By the Late Bronze Age 2950 to 2875 years ago, the EEF ancestry in southern British people surged to 38%. Using a mathematical model, the team calculated that by the Iron Age, starting 2750 years ago, EEF were the source of about half of the DNA of people living in southern Britain. "We're seeing people moving on a scale to sufficiently alter the genetic makeup of the populations in southern Britain," Armit says.

The researchers found that the genomes of the migrants buried near Dover were closely related to those of people then living at sites in France and Spain, including skeletons tied to the Urnfield culture of Central Europe, thought to have links to early Celtic languages. But Reich says scientists need more DNA from Europe, especially France, to pinpoint the migrants' homeland.

The results boost a theory that Celtic languages spread from France to Britain in the Late Bronze Age. "I'm delighted, but I'm obviously biased," says philologist Patrick Sims-Williams at Aberystwyth University, who backs this idea. But some modern people who identify as "Celts" (a term archaeologists say includes diverse material cultures) and speak Celtic languages they consider indigenous, such as Irish, Scottish Gaelic, or Welsh, could find the results "unsettling," Sims-Williams says. "Celtic becomes just one of a succession of languages that migrants have brought to Britain over the course of 2000 years or so: Latin with the Romans, English with the Anglo-Saxons, Norse with the Vikings, French with the Normans."

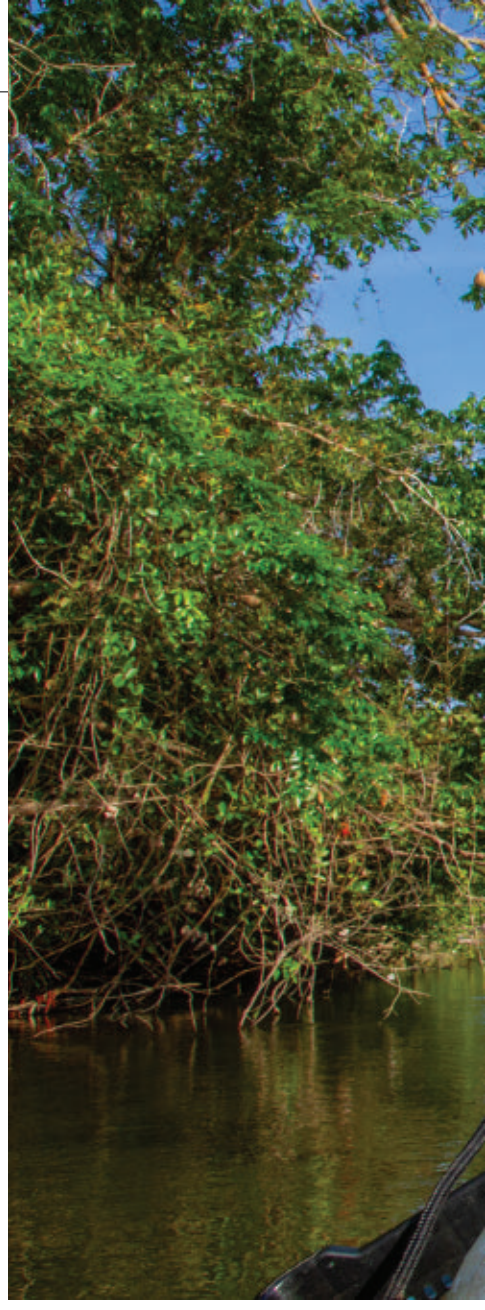
The later migrations and political dominance of the English eventually marginalized Celtic languages, leaving them spoken chiefly beyond England's borders, Armit says.

Others agree that the new study fits historical and archaeological evidence of close ties between Britain and Europe in the Late Bronze Age, when sailors used the white cliffs of Dover as a guide to cross the English Channel at its narrowest point and find the entrance to the Thames River. The new study shows "centuries of migration, with men, women, and children across all levels of society moving," says University of York archaeologist Lindsey Büster—the kind of population shift that sustains language change. By the end of the Bronze Age, cultures on both sides of the channel had many similarities.

But tying a language to DNA is tricky, cautions geneticist Dan Bradley of Trinity College Dublin. He's seeking traces of French blood in ancient Irish DNA, to find out whether the early Celtic speakers also made it to Ireland, today's Celtic heartland. ■

FEATURES





AMAZON DELIVERY

A rainforest expedition brings COVID-19 vaccines to some of Brazil's most remote and vulnerable groups

Story and photography by **Sofia Moutinho**

"Vaccine! Let's vaccinate everyone!" Shouts go up from the steep and muddy banks of the Iriri River in the heart of the Brazilian Amazon forest. A team of doctors, nurses, and nongovernmental organization (NGO) workers are hauling plastic foam boxes from a small aluminum motorboat. Dozens of residents of this village, Boa Esperança, gather on the balcony of a wooden house to avoid the scorching Sun and await their COVID-19 vaccines.

Medical teams are a rare sight here. This crew has traveled 9 hours in pickup trucks along bumpy roads from the closest city of Altamira, then sailed a few more hours. They are at the start of a 17-day river trip to give shots at settlements in preserved areas of the world's biggest tropical forest.

Most residents here are riverine people (*ribeirinhos* in Portuguese)—riverside dwellers who have lived in the Amazon for centuries, relying on artisanal fishing and harvesting. The Brazilian government views them as a traditional group—a category that also includes Indigenous tribes and descendants of runaway enslaved Africans, known as *quilombolas*. These groups were given priority status to receive the COVID-19 vaccine in a national plan released in December 2020, following pressure from human rights activists and local leaders. Many Indigenous groups, which receive care from a special federally funded health department, have been vaccinated. But riverine peoples depend on local municipalities, which are mostly too poor to do much, for their health care. As a result, whereas city dwellers in Brazil are already getting boosters, many people living on remote riverbanks still wait in isolation for their first or second shots.

"It is a population very difficult to reach that has suffered from neglect from the state for centuries," says the doctor in charge of the expedition, Erika Pellegrino of the Federal University of Pará, Altamira. The municipal health department of Altamira supplied the COVID-19 shots—two-dose AstraZeneca and single-dose Janssen vaccines—that the team is delivering along with various childhood vaccines. But NGOs, including the local Socio-environmental Institute and the U.S.-based group Health in Harmony, organized and funded the expedition itself, which cost nearly $20,000. Grassroots associations helped with the infrastructure, converting houses into improvised vaccination centers.



In small aluminum boats (left) a crew transports COVID-19 vaccines through the shallow Riozinho do Anfrísio River.

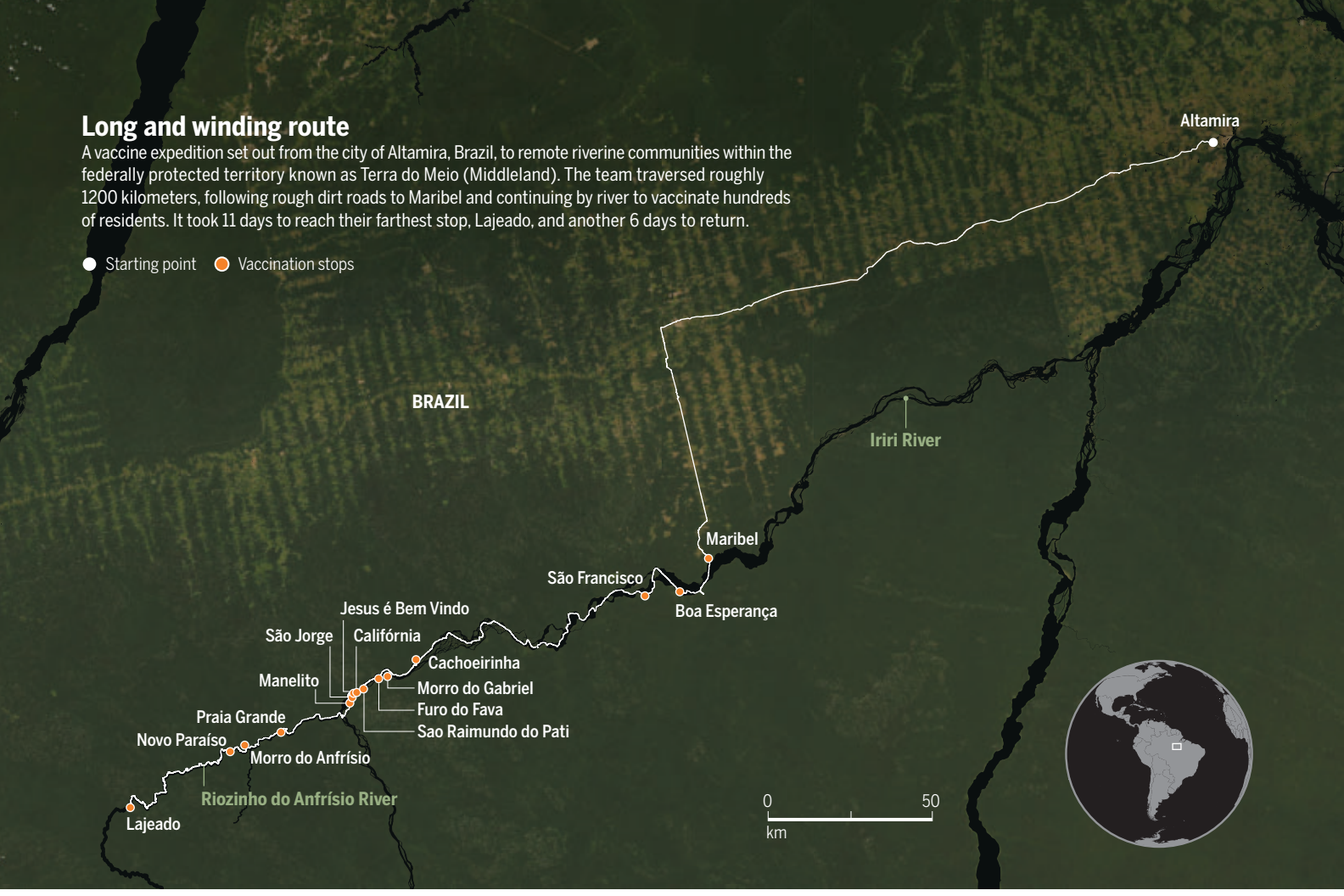
It has taken four long trips between June and September to vaccinate more than 700 people living on three reserves on the banks of the Iriri, Xingu, and Riozinho do Anfrísio rivers. The number of people vaccinated so far might seem small, but it includes many of the estimated 1700 riverine people living in these reserves, located in the vast mosaic of federally protected lands called Terra do Meio (Middleland)—an area about the size of Iceland that abuts Indigenous reserves. The whole Amazon forest likely contains tens of thousands of riverine people, but they haven't been formally counted. Also lacking are centralized data on how many have been vaccinated through government campaigns or NGO missions like this one.

"It is a relatively small number of people spread on a gigantic territory," Pellegrino says of the riverine groups. "And that is our biggest challenge."

"HELLO, HELLO EVERYONE!" The expedition is coming in 2 days with COVID-19 vaccines and exams. Copy that?" a crew member in Boa Esperança announces into a ham radio transmitter. "Positive! Copy that!" come the crackling responses. As the vaccination team proceeds up the river, it relies on radios like this, the only form of communication available to many riverine people, to spread the word about upcoming visits. That's how 63-year-old Francisco dos Santos, a nut harvester, fisherman, and local singer known as Chico Preto, learned about the vaccination.

Dos Santos's 79-year-old wife, Maria Madalena Freire, narrowly survived a severe case of COVID-19 in August 2020. Local organizations and NGOs offered her an emergency air evacuation, which she declined, saying she wished to "die in the forest." But that experience impelled the couple to sail for 2 hours in a motorized canoe to a vaccination site in the village of Manelito, about 160 kilometers up the river from Boa Esperança. "We wish we could have brought more people with us, but the fuel was not enough," dos Santos says. The medications, checkups, and basic blood tests offered by the team were an extra draw, he says.

Dos Santos, like most riverine people in the region, is a descendant of Indigenous people and "rubber soldiers," immigrants from northeastern Brazil sent by the government to the forest to harvest latex for the U.S. war effort during World War II. They now support themselves by fishing, hunting, and harvesting Brazil nuts and the



coconutlike fruits of the babaçu palm. None has access to running water, and only a few have electricity from solar panels or diesel generators for a few hours a day. Their communities have no doctors and are days away from a hospital. Four nurse technicians, hired by the municipality with support from NGOs during the pandemic, are the only medical support for kilometers. They are supposed to assist all families living in the reserves in Terra do Meio. "God knows how difficult it is to get medical assistance here," Freire says. "When a health team comes, I am the first to run after them."

It's a dangerous environment for a COVID-19 outbreak. "The vulnerability [of these groups] is socio-epidemiologic," says Douglas Rodrigues, a physician at the Indigenous clinic at São Paulo State University's hospital, who has been working with Indigenous and riverine communities for more than 30 years. Along with a lack of proper health care, the riverine people have elevated rates of comorbidities such as diabetes and infectious disease, which raise their risk of severe disease and death from the virus, he says. Their communal lifestyle, in which large families often share a one-room house, also makes it easier for SARS-CoV-2 to spread. Isolating the infected "is just not possible in their way of life," Rodrigues says.

Riverine and Indigenous communities have restricted visitors to their territories since the start of the pandemic to prevent the disease's arrival. But illegal activities such as mining, logging, and land grabbing, which have spiked in the past year, continue to bring in outsiders. As a result, COVID-19 has spread with devastating consequences.

Among Indigenous peoples in the region, disease incidence and mortality during the pandemic's first wave in 2020 were respectively 136% and 110% higher than the national average, according to a study published in April in the journal *Frontiers*. There are no official numbers for riverine peoples, as they are not tracked separately by the Ministry of Health.

TO GET TO the expedition's farthest stop, the village of Lajeado in the Riozinho harvesting reserve, the vaccination crew must travel more than 600 kilometers up the narrow Riozionho do Anfrísio. Some parts of the river look like a swamp, with shallow, dark water full of caimans. Others are studded with rocks, rapids, and little waterfalls. Only small boats called *rabetas*—with a retractable propeller attached to a long shaft—can navigate these shallow waters.

"Hold on!" shouts one of the boats' captains while lifting the propeller from the water to avoid a fallen tree on the river's bottom. The co-captain stands on the prow using a machete to chop away overhanging tree limbs. Crew members duck as spiky vegetation along the shore sweeps past the boat.

Aboard are six people and precious cargo: a box containing dozens of vials of the AstraZeneca vaccine and a few single-dose Janssen syringes. They must be kept between 2°C and 8°C, and are carefully stored on ice that is supposed to last 5 days. By then, the team will have returned to less treacherous waters, and can transfer to a larger boat with a solar-powered freezer.

By the time the team has finished the 2-day-long journey from its previous vaccination stop in Morro do Anfrísio to Lajeado, one of the *rabetas* has broken down three times. Arriving just before sunset, the crew set up their hammocks. Vaccinations will start early in the morning. A nurse technician jokes that she's having withdrawal symptoms from going so many hours without vaccinating anyone. "My hands are shaking to give these shots!" she says.

GRAPHIC: K. FRANKLIN/*SCIENCE*

People on the neighboring Indigenous lands of Xipaya and Kuruaya have been vaccinated for months, thanks to the federally funded health department that serves them. But many of the municipalities responsible for riverine peoples have neglected vaccinations. Altamira started to make the vaccine available to these groups in May, but only after the federal public attorney's office started to ask municipalities to enforce their priority in the national rollout.

Such disparities are nothing new, says anthropologist Roberta Cerri at the Oswaldo Cruz Foundation, who is researching health access in these communities. She believes Brazil should have a separate health strategy for all traditional people living in the Amazon forest. "Neither COVID-19 nor any other infectious disease differentiates if a person is from an Indigenous tribe or a riverine community," she says.

Consistent vaccination across both Indigenous and riverine communities is also key to safeguarding uncontacted tribes in their vicinity. There are about 100 of these groups in the forest, according to the Brazilian government's National Indian Foundation, and they are especially vulnerable to many infectious diseases. "For isolated groups, a flu virus is already a pandemic, to say nothing of the coronavirus," Rodrigues says. In keeping with Brazil's long-standing no-contact policy, they won't be vaccinated. But inoculating their neighbors creates a safety barrier, reducing the chances of the disease reaching them.





A nurse technician (top) fills a syringe with COVID-19 vaccine in Maribel, Brazil. Nut harvester Francisco dos Santos shows his vaccination card at the Manelito vaccination stop.

PHOTOS: SOFIA MOUTINHO

IN LAJEADO, the crew improvises a vaccination post with school chairs and a table under the bright yellow flowers of an *ipê-amerelo* tree. After a full day of vaccinations, only one woman is left in line. But opening a vaccine vial just for her would mean wasting four of its five doses, which expire 6 hours after opening. Samya Mauad, a nurse technician, quickly finds three more villagers willing to get shots. Still missing one person, she tries to recruit a man who came to vaccinate his toddler against childhood diseases. But he refuses.

"Did you ever get vaccinated?" she asks. He nods. "So why don't you take this one? The more people get vaccinated, the more protected we get." Villagers watching the scene jump in with words of encouragement. But the pressure irritates the man, and he walks away, carrying his daughter.

It's not the first time Mauad has encountered hesitancy. Most often, people are afraid of side effects, and she manages to convince them, she says. But even in an area without cellphone reception and with only a few internet hubs, fake news spreads. People have told Mauad that the injection contains a "devil's chip" to control minds, that it turns people into homosexuals, or that it makes them develop AIDS.

Some of the misinformation may come from certain evangelical missionaries active in the region, who are currently under scrutiny by a parliamentary commission investigating antivaccine propaganda and human rights violations during the pandemic. Other falsehoods come from the highest levels of the Brazilian government: President Jair Bolsonaro, who regularly criticizes COVID-19 vaccines, said on a Facebook live broadcast in October that citizens were developing AIDS after receiving the shot. Few officials have countered his message. "The Brazilian federal government has made no effort to instruct people about the importance of vaccinating," Rodrigues says.

Still, the crew estimates that fewer than 1% of the people approached on this trip have refused shots. The team is now planning trips to bring dentists and doctors to the region and has started to organize an expedition to vaccinate teenagers, who have been eligible for vaccination since August.

Despite their isolation, many riverine people know the pandemic is causing global waves of illness and want to keep COVID-19 at bay. Antonio Silva Matos, a gold miner in the city of Novo Progresso traveled about 300 kilometers back to Lajeado, near his home, to get his shot. (Health authorities in Novo Progresso would only vaccinate residents, not migrant laborers, he says.) The trip required a pickup truck, a bus, a motorbike, and a motorboat—and cost him 3000 reais (about $600). He didn't regret it. "So many people have the vaccine at their doors and don't want it," he says. "I just don't get it." ■

INSIGHTS

PERSPECTIVES



Ichthyosaurs evolved in Triassic oceans where common prey items were shelled cephalopods, such as ammonoids.

ILLUSTRATION: STEPHANIE ABRAMOWICZ/DINOSAUR INSTITUTE, NATURAL HISTORY MUSEUM OF LOS ANGELES COUNTY

EVOLUTION

Early and fast rise of Mesozoic ocean giants

A whale-sized ichthyosaur shows how fast these reptiles evolved

By **Lene Liebe Delsett**[1,2] *and* **Nicholas D. Pyenson**[1,3]

We live in a time of ocean giants: Today's blue whales, right whales, sperm whales, and killer whales are not just extremely large marine predators, they each represent the largest-sized examples ever for their respective lineages (*1*, *2*). Notably, whale gigantism appears to be a relatively recent phenomenon, only having evolved in the past few million years, which is but a fraction of their total evolutionary history (*3*). Does this delayed rise of gigantism hold true for ocean giants at other times in the past? After all, dozens of reptile and mammal lineages have invaded ocean ecosystems for the past 300 million years (*4*), and whale-sized vertebrates have evolved many times (*5*). On page 1578 of this issue, Sander *et al.* (*6*) describe a very large species of ichthyosaur, *Cymbospondylus youngorum* sp. nov., from early Middle Triassic rocks of Nevada that appears to be the first ocean giant.

Ichthyosaurs were among the first vertebrates to invade the oceans after the Permian-Triassic mass extinction (252 million years ago), and they persisted through most of the Mesozoic era. The specimen described by Sander *et al.* is well preserved and very large. The skull alone is about the size of a grand piano, which places total length estimates for *C. youngorum* at more than 18 m long, similar to that of an adult fin whale. Even more interesting is the age of the species—it evolved about 2.5 million years after the appearance of *Cartorhynchus*, the oldest proposed ichthyosaur relative from China, which is no bigger than a laptop computer (*7*). Thus, the early evolutionary history of ichthyosaurs seems to show a rapid rise of gigantism, with *C. youngorum* representing the largest vertebrate of its time, either at sea or on land.

Previous discoveries from Early Triassic rocks of China have established that the first ichthyosaurs were relatively small (*7*), but Early Triassic *Cymbospondylus* fossils from Svalbard (an archipelago in the Arctic) (*8*) have shown that large bodies evolved in the earliest phases of ichthyosaur macroevolution. Fossils like these and others prompted an "early burst" evolutionary model in the wake of the Permian-Triassic mass extinction (*9*). In this context, Sander *et al.*'s discovery of *C. youngorum* provides a new size maximum that is notably close to the time origin for the oldest ichthyosaurs, compressing the timeline of earlier models into just the first few million years of ichthyosaur history (see the figure).

The speed of ichthyosaur gigantism might be unparalleled among marine vertebrates. One of the major contributions of Sander *et al.* is the comparison of this rapid increase in size with comparative data from the evolutionary history of whales, which are the closest ecomorphological comparatives of ichthyosaurs. Whereas it took whales about 90% of their 55-million-year history to evolve into giants, ichthyosaurs—as exemplified by *C. youngorum*—did it in the first 1% of their 150-million-year history.

There is an ongoing debate about how fast ecosystems can recover from mass extinctions. In the Middle Triassic, marine ecosystems were still recovering after the Permian-Triassic mass extinction, an event during which more than 80% of marine genera died out. Ichthyosaurs were the first vertebrates to achieve dominance in the top spots of ocean food webs, but the earliest phases of their transition from land to sea are still obscure (*4*), unlike that of whales (*10*). Today's whales and other ocean giants take advantage of short food chains to access high densities of zooplankton (*1*) and are themselves vital contributors to nutrient cycling (*11*). However, this configuration

[1]Department of Paleobiology, National Museum of Natural History, Smithsonian Institution, Washington, DC, USA. [2]Natural History Museum, University of Oslo, Oslo, Norway. [3]Burke Museum of Natural History and Culture, Seattle, WA, USA. Email delsettl@si.edu; pyensonn@si.edu

might have been different in the Early and Middle Triassic (*12*). If ocean productivity was crucial for fueling the rise of today's giants (*3*), was it not for ichthyosaurs, too?

Sander *et al.* provide insight into how Triassic ocean ecosystems promoted and maintained vertebrate gigantism. *C. youngorum* was one of several species of ichthyosaurs recovered from the same strata of rocks in Nevada. The fossil invertebrate fauna from the site is dominated by ammonoids—now-extinct shelled cephalopods—which represented the likeliest diet for ichthyosaurs. By modeling the energy flux through this ecosystem, Sander *et al.* show that there was enough food in the open ocean to support large predators and that the marine food web was stable, populated by pelagic conodonts (extinct, eel-like jawless vertebrates), ammonoids, and small ichthyosaurs as prey for larger predators. One possibility that should be explored further is whether conditions in the open oceans in the Proto-Pacific were better for promoting gigantism than the hotter, shallower oceans in the Tethys Sea, where the Chinese early ichthyosaurs are found.

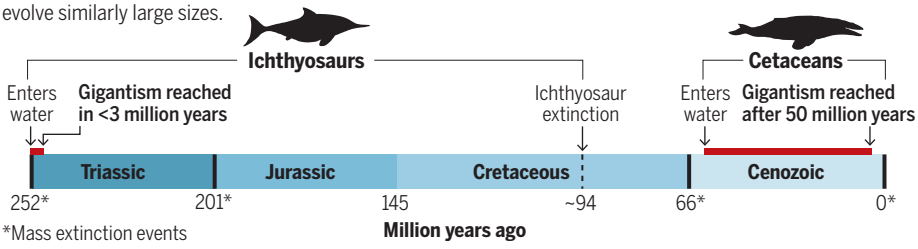
Regardless of pace, do ocean giants share similar pathways to gigantism? The comparison of ichthyosaurs to whales leads to even more questions about the patterns and processes of these parallel evolutionary innovations: When and how did ichthyosaurs evolve key adaptations for live birth underwater, a stable body temperature, crucial feeding strategies, and new ways to swim? We are used to thinking of ichthyosaurs as the dolphins of the Mesozoic, but that might cloud important and interesting differences. For example, Triassic ichthyosaurs such as *Cymbospondylus* had longer necks, torsos, and tails and were probably more "lizard-like," with different foraging and behavioral ecologies compared with the more iconic ichthyosaurs from the Jurassic (*13*).

The largest whales today are threatened with extinction in large part because of their extreme size, which makes them dependent on ephemeral food resources, and because of human activities. For ichthyosaurs, the early innovation of gigantism may also have carried extinction risk, but the history of body size in this group is complex. Ichthyosaurs were top ocean predators for three times as long as whales have even existed. The group survived the Triassic-Jurassic mass extinction event, which reduced their diversity in size and feeding ecology, 50 million years after the appearance of *C. youngorum* (*13*). Ichthyosaurs then survived more than 100 million years longer until finally going extinct in the Cretaceous period, possibly because of climate change (*14*).

It remains unclear whether ichthyosaurs were ecosystem engineers in the way that some large species of whales are (*11*), but it is clear that sunken ichthyosaur carcasses served as a substrate for seafloor communities in the way that whalefalls do today (*15*). Ichthyosaur history tells us that ocean giants are not guaranteed features of marine ecosystems, which is a valuable lesson for all of us in the Anthropocene, especially if we want to sustain the presence of the surviving ocean giants among us that contribute to our own well-being (*11*). ■

Two stories of ocean gigantism, eras apart

Ichthyosaurs were the first ocean giants, and the discovery of *Cymbospondylus youngorum* narrows the window for how fast they evolved extremely large body sizes. Although ichthyosaurs evolved whalelike sizes in the first 1% of their evolutionary history, it later took cetaceans nearly 90% of their evolutionary history to evolve similarly large sizes.



GRAPHIC: K. FRANKLIN/*SCIENCE*

REFERENCES AND NOTES

1. J. A. Goldbogen *et al.*, *Science* **366**, 1367 (2019).
2. N. D. Pyenson, G. J. Vermeij, *Biol. Lett.* **12**, 20160186 (2016).
3. G. J. Slater, J. A. Goldbogen, N. D. Pyenson, *Proc. Biol. Sci.* **284**, 20170546 (2017).
4. R. Motani, G. J. Vermeij, *Biol. Rev. Camb. Philos. Soc.* **96**, 1769 (2021).
5. N. P. Kelley, N. D. Pyenson, *Science* **348**, eaaa3716 (2015).
6. P. M. Sander *et al.*, *Science* **374**, eabf5787 (2021).
7. R. Motani *et al.*, *Nature* **517**, 485 (2015).
8. V. S. Engelschiøn *et al.*, *Norwegian Journal of Geology* **98**, 239 (2018).
9. B. C. Moon, T. L. Stubbs, *Commun. Biol.* **3**, 68 (2020).
10. N. D. Pyenson, *Curr. Biol.* **27**, R558 (2017).
11. M. S. Savoca *et al.*, *Nature* **599**, 85 (2021).
12. C. P. A. Smith *et al.*, *Sci. Rep.* **11**, 19657 (2021).
13. P. M. Thorne, M. Ruta, M. J. Benton, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 8339 (2011).
14. V. Fischer, N. Bardet, R. B. J. Benson, M. S. Arkhangelsky, M. Friedman, *Nat. Commun.* **7**, 10825 (2016).
15. S. Danise, R. J. Twitchett, K. Matts, *Nat. Commun.* **5**, 4789 (2014).

10.1126/science.abm3751

MAGNETISM

Magnetic control in the terahertz

Strong magnon-phonon coupling may help develop superfast optical drives

By **Dominik M. Juraschek**[1,2] *and* **Prineha Narang**[1]

Magnetic materials are central to information storage devices, with ongoing research seeking to develop faster and more energy-efficient systems. The individual bits of data, written as 1's and 0's, can be stored as different orientations of magnetic moments. Conventionally, an electromagnetic head is used to flip these bits between 1 and 0, and the read and write speeds using this method are currently limited to gigahertz frequencies. A substantial amount of energy is used to drive electric currents that generate the magnetic fields. Instead of manipulating magnetic moments with slow or static magnetic fields, an intriguing alternative involves coupling light from a laser to their quantum mechanical states. On page 1608 of this issue, Mashkovich *et al.* (*1*) report the coupling of light, magnetism, and the crystal structure of a material, which opens new possibilities for controlling the magnetic state of materials on very short time scales.

Over the past two decades, researchers have explored the use of laser pulses to modify magnetism and push the boundaries of read-write speeds to the terahertz and even petahertz range. When a laser is shone onto a magnetic material, the electromagnetic field of the incident light interacts with the spins of the atoms and can trigger collective precessions of the spins around their equilibrium orientation. These precessions form quantized quasiparticles, collective interactions of atoms that act as single particles, that are called "magnons." If the amplitude of these precessions becomes large enough, entire magnetic domains in the material may change their orientation,

[1]Harvard John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA. [2]School of Physics and Astronomy, Tel Aviv University, Tel Aviv 69978, Israel. Email: djuraschek@tauex.tau.ac.il; prineha@seas.harvard.edu

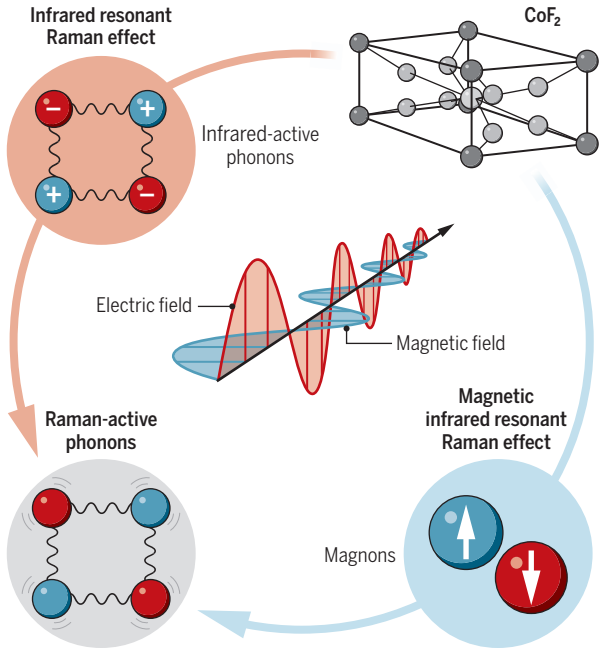
essentially flipping a bit (*2*). To fully realize this optical control of magnetism, scientists need to unravel the fundamental physical processes governing the transfer of energy from the laser pulse to electrons in the material. Recently, researchers have found ways to manipulate magnetic orders inside materials indirectly. Instead of directly exciting magnons in a material, these methods couple to magnetic order by manipulating the crystal structure of the material with light (*3*).

The magnetic order of a material depends on the geometry and symmetry of its crystal lattice, which affects the distances between the atoms and thus the quantum-mechanical exchange interactions that are responsible for magnetic ordering. Therefore, a natural approach is to look at collective vibrations of atoms, because they are directly linked to changes in the crystal structure and subsequently to the desired modulations of the exchange interactions. These crystal lattice vibrations follow defined patterns of atomic motion, and their quantized quasiparticles are called "phonons." Just like the photons from a laser, phonons can interact with magnetic order in a way similar to their luminous siblings.

The spin-lattice interaction makes it possible to use phonons as the intermediary between the laser pulse and the magnetic order. However, efficiently transferring energy from photons to phonons has been a long-standing challenge, because phonon frequencies often lie in an electromagnetic spectrum known as the "terahertz gap" between 0.3 and 30 THz, where there is a lack of practical technologies for generating and detecting electromagnetic waves. This has been overcome in recent years, and the transfer of energy from phonons to magnons has been achieved experimentally and theoretically (*4*, *5*). As powerful tabletop terahertz sources are becoming more widely accessible, more demonstrations of optically activated spin-lattice interactions have been reported, such as those that show the reorientations of magnetic domains or the generation of magnetization in antiferromagnets (*6*–*8*). In these studies, the transfer of energy is directed from structural to magnetic order, but there is no fundamental reason why this could not go the other way round as well. Because photon-magnon coupling tends to be weaker than photon-phonon coupling, a photon-to-magnon-to-phonon system would require a more efficient coupling mechanism between the different quasiparticles.

Two pathways to the Raman effect

In Raman scattering, the electric field of the incident light is coupled to infrared-active phonons, which convert the energy into lattice vibrations. In the magnetic analog of the Raman effect, the magnetic field of the incident light is coupled to magnons in the material and generates lattice vibrations in a similar fashion.



Mashkovich *et al.* demonstrate an efficient coupling between magnons and phonons in the prototype antiferromagnetic compound cobalt fluoride (CoF_2). They identify a magnon that couples its spin precession to the magnetic field of the light from a laser pulse, which enables one to drive the precession resonantly and with a large amplitude. This observation was linked to coherent structural oscillations in the crystal corresponding to a phonon mode, which in turn is coupled to the electron orbits and the magnetization of the bulk material (*6*, *9*). However, despite the precessing and oscillating of the magnon and phonon at the same time, there are still multiple possible channels through which the phonon could have received its energy. How can one be sure that the phonon is not being driven by the light directly? To answer this, Mashkovich *et al.* performed temperature-dependence measurements of the excitation mechanism. They found that the oscillation of the phonon vanishes at the same temperature at which the material loses its magnetic order, which strongly suggests that the phonon receives most of its energy from the magnetic system, through a previously unknown energy-transfer mechanism and with unprecedented efficiency.

From a theoretical perspective, we note that the coupling of light, magnons, and phonons demonstrated by Mashkovich *et al.* is reminiscent of that recently predicted in the infrared resonant Raman effect for a photon-to-phonon-to-phonon system (*10*), in which the light's electric field and the phonon's electric dipole moment take the place of the light's magnetic field and the magnon's magnetic moment. Reciprocally, the coupling demonstrated by Mashkovich *et al.* could be used to generate a magnetic infrared resonant Raman effect, which adds yet another route to controlling magnetism through the crystal lattice (see the figure).

Research on phonon-driven magnetic systems, also dubbed phonomagnetism or magnetophononics, has seen a rapid increase in recent years, and this trend will likely accelerate as more laboratories with strong terahertz sources become operational. Theoretical and computational work plays a pivotal role in furthering the fundamental understanding and description of these emerging phenomena, not only reactively but also proactively in predicting new physical mechanisms and materials in which to find them (*11*–*15*).

A central challenge for the research field is finding compounds with strong spin-lattice couplings that can help achieve terahertz-speed magnetic control. Here, first-principles calculations combined with high-throughput methods and artificial intelligence–driven techniques may be the solution, because these methods can scan over a vast range of possible candidate compounds. The study of Mashkovich *et al.* has set a precedent for where to start the search. ■

REFERENCES AND NOTES

1. E. A. Mashkovich *et al.*, *Science* **374**, 1608 (2021).
2. S. Schlauderer *et al.*, *Nature* **569**, 383 (2019).
3. A. S. Disa, T. F. Nova, A. Cavalleri, *Nat. Phys.* **17**, 1087 (2021).
4. T. F. Nova *et al.*, *Nat. Phys.* **13**, 132 (2017).
5. D. M. Juraschek, M. Fechner, A. V. Balatsky, N. A. Spaldin, *Phys. Rev. Mater.* **1**, 014401 (2017).
6. A. S. Disa *et al.*, *Nat. Phys.* **16**, 937 (2020).
7. D. Afanasiev *et al.*, *Nat. Mater.* **20**, 607 (2021).
8. A. Stupakiewicz *et al.*, *Nat. Phys.* **17**, 489 (2021).
9. P. G. Radaelli, *Phys. Rev. B* **97**, 085145 (2018).
10. G. Khalsa, N. Benedek, J. Moses, *Phys. Rev. X* **11**, 021067 (2021).
11. M. Gu, J. M. Rondinelli, *Phys. Rev. B* **98**, 024102 (2018).
12. M. Fechner *et al.*, *Phys. Rev. Mater.* **2**, 064401 (2018).
13. M. Rodriguez-Vega *et al.*, *Phys. Rev. B* **102**, 081117(R) (2020).
14. D. M. Juraschek, P. Narang, N. A. Spaldin, *Phys. Rev. Res.* **2**, 043035 (2020).
15. D. M. Juraschek, D. S. Wang, P. Narang, *Phys. Rev. B* **103**, 094407 (2021).

10.1126/science.abm0085

GRAPHIC: A. MASTIN/*SCIENCE*

CORONAVIRUS

A new tool to probe SARS-CoV-2 variants

Virus-like particles offer a new way to investigate genetic variation in SARS-CoV-2

By **Bryan A. Johnson**[1] *and* **Vineet D. Menachery**[1,2,3,4]

Although efforts have been made to understand the biology of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a major focus has been on investigating genetic variation in the virus. However, progress is hampered by the need to perform experiments involving SARS-CoV-2 in biosafety level 3 (BSL3) laboratories, which require substantial training for safe operation. On page 1626 of this issue, Syed *et al.* (*1*) offer an alternative to using live virus, introducing a new SARS-CoV-2 virus-like particle (VLP) system. The authors innovate on previous VLP systems by incorporating a reporter construct to study infection (*1, 2*). Illustrating the system's utility, they use VLPs to characterize mutations in SARS-CoV-2 variants of concern.

SARS-CoV-2 VLPs are created by expressing the four structural proteins—spike, membrane, envelope, and nucleocapsid—in a packaging cell line (*2*). Upon expression, VLPs consisting of these four proteins and a lipid membrane self-assemble and are released from the cell (*3*). Despite resembling SARS-CoV-2 morphologically, traditional VLPs cannot be used to study the effect of a mutation on fitness because they lack genetic material to deliver to target cells (*2*). Syed *et al.* introduced a key innovation. They first identified the SARS-CoV-2 packaging signal, a genetic marker used to identify full-length genomes for packaging into the virion (*4*). This packaging signal was incorporated into the 3' untranslated region of a luciferase reporter plasmid, causing the resulting transcripts to be packaged within VLPs. Syed *et al.* show that VLPs deliver these luciferase reporters to target cells, allowing the resulting signal to be used as a proxy for SARS-CoV-2 infection. Thus, the effects of particular mutations on the strength of the luciferase signal can be used to determine modulation of SARS-CoV-2 infection (see the figure).

In the broader context of studying SARS-CoV-2 genetic variation, VLPs represent a middle ground between two commonly used methodologies: infectious clones and pseudovirus vectors. SARS-CoV-2 infectious clones are the gold standard because they create recombinant virus, incorporating mutations anywhere in the genome (*5*). However, using SARS-CoV-2 infectious clones is technically challenging and creates live SARS-CoV-2 that requires BSL3 laboratories for study. This limits the use of SARS-CoV-2 infectious clones to laboratories with access to such facilities and willingness to invest in developing a specialized skill set.

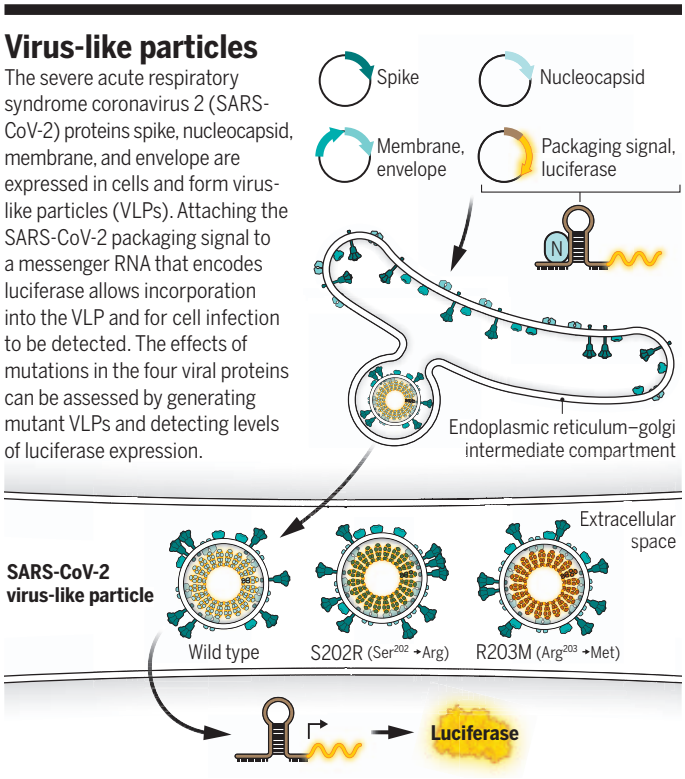
Pseudovirus systems are the leading alternative to using SARS-CoV-2 infectious clones. In these systems, SARS-CoV-2 spike protein is expressed in cells along with a noncoronavirus packaging system and a reporter gene, with the most common being lentivirus-based (*6*). Like the VLPs developed by Syed *et al.*, pseudoviruses self-assemble, incorporating spike proteins on their surface and packaging reporter messenger RNA (mRNA) (*6*). The primary advantage of pseudovirus systems is their ease of use, allowing rapid analysis of spike mutations. Pseudoviruses can be generated in the widely available 293T cell line by simply expressing a small number of proteins (*6*). Additionally, because pseudoviruses replace replication genes, they do not undergo continued amplification in target cell lines (*6*). This makes them safe to use in BSL2 laboratories, which are available to most researchers. However, the only SARS-CoV-2 protein incorporated into pseudoviruses is spike. Because substantial genetic variation occurs outside of spike, the pseudovirus systems have limited applicability to study SARS-CoV-2 variants.

The SARS-CoV-2 VLPs used by Syed *et al.* offer researchers several advantages over pseudoviruses. Rather than relying on the packaging machinery of another virus, VLPs use SARS-CoV-2 proteins and recapitulate packaging, assembly, and release, as occurs in genuine virus infection (*3*). In principle, this allows the effects of variant mutations on these processes to be studied. Similarly, because all four structural proteins are incorporated into SARS-CoV-2 VLPs, additional genetic variation can be captured. Like pseudoviruses, VLPs do not undergo subsequent rounds of replication, allowing them to be used safely in BSL2 laboratories.

Illustrating the utility of SARS-CoV-2 VLPs, Syed *et al.* characterized several nucleocapsid mutations. SARS-CoV-2 nucleocapsid is a hotspot for coding mutations, particularly within its serine-rich (SR) motif (*7, 8*). Although its exact function is unclear, the SR motif has many phosphorylated amino acids and is located within a region of intrinsic structural disorder (*7, 9, 10*). Using their SARS-CoV-2 VLP system, Syed *et al.* analyzed the effects of several common nucleocapsid

Virus-like particles

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) proteins spike, nucleocapsid, membrane, and envelope are expressed in cells and form virus-like particles (VLPs). Attaching the SARS-CoV-2 packaging signal to a messenger RNA that encodes luciferase allows incorporation into the VLP and for cell infection to be detected. The effects of mutations in the four viral proteins can be assessed by generating mutant VLPs and detecting levels of luciferase expression.



GRAPHIC: V. ALTOUNIAN/*SCIENCE*

[1]Department of Microbiology and Immunology, University of Texas Medical Branch, Galveston, TX, USA. [2]Institute for Human Infection and Immunity, University of Texas Medical Branch, Galveston, TX, USA. [3]World Reference Center of Emerging Viruses and Arboviruses, University of Texas Medical Branch, Galveston, TX, USA. [4]Center for Biodefense and Emerging Infectious Diseases, University of Texas Medical Branch, Galveston, TX, USA. Email:Vimenach@utmb.edu

mutations and found that several enhanced infection, including those present in the Alpha, Gamma, and Delta variants (*7*). These data are consistent with findings using SARS-CoV-2 infectious clones (*1, 11*).

The finding that nucleocapsid mutations enhance SARS-CoV-2 infection has important implications. To date, most studies of SARS-CoV-2 genetic variation have focused on spike (*12*). This is understandable, because spike binds to the host cell receptor angiotensin-converting enzyme 2 (ACE2), and is thus the primary determinant of infection (*13*). Additionally, because spike is the target of available vaccines, determining if mutations affect protection is a pressing question (*12*). However, recent studies suggest that nucleocapsid mutations lead to enhanced virulence and fitness, highlighting the need to characterize genetic variation elsewhere in the viral genome (*11*). Because SARS-CoV-2 VLPs recapitulate enhancement of infection by these nucleocapsid mutations, they can be used to characterize mutations in emerging variants, such as deletion of amino acids 31 to 33 in the nucleocapsid protein of the Omicron variant.

Although a promising platform, there are limitations of this SARS-CoV-2 VLP system. Only the four structural proteins are present. Thus, like pseudoviruses, the scope of variation that can be captured is limited. For example, variant mutations in the viral replication machinery cannot be examined with VLPs (*14*). Additionally, while allowing for safe use in BSL2 laboratories, the inability of VLPs to undergo continued replication makes them unsuitable to study virulence or transmission. Furthermore, although data presented by Syed *et al.* suggest that enhancement of infection by VLPs and live SARS-CoV-2 are correlated, additional work is needed to determine how closely VLPs model infection. As SARS-CoV-2 evolves, it is critical that the effects of new mutations are characterized. ■

REFERENCES AND NOTES

1. A. M. Syed *et al.*, *Science* **374**, 1626 (2021).
2. H. Swann *et al.*, *Sci. Rep.* **10**, 21877 (2020).
3. P. S. Masters, *Adv. Virus Res.* **66**, 193 (2006).
4. P.-K. Hsieh *et al.*, *J. Virol.* **79**, 13848 (2005).
5. X. Xie *et al.*, *Cell Host Microbe* **27**, 841 (2020).
6. M. Chen, X.-E. Zhang, *Int. J. Biol. Sci.* **17**, 1574 (2021).
7. Q. Ye *et al.*, *Protein Sci.* **29**, 1890 (2020).
8. J. A. Plante *et al.*, *Cell Host Microbe* **29**, 508 (2021).
9. M. Bouhaddou *et al.*, *Cell* **182**, 685 (2020).
10. A. D. Davidson *et al.*, *Genome Med.* **12**, 68 (2020).
11. B. A. Johnson *et al.*, *bioRxiv* 10.1101/2021.10.14.464390 (2021).
12. W. T. Harvey *et al.*, *Nat. Rev. Microbiol.* **19**, 409 (2021).
13. C. B. Jackson *et al.*, *Nat. Rev. Mol. Cell Biol.* (2021).
14. J. L. Mullen *et al.*, https://outbreak.info/ (2020).

ACKNOWLEDGMENTS

V.D.M. is funded by the National Institutes of Health and National Institute of Allergy and Infectious Diseases (grants AI153602, 1R21AI145400, and R24AI120942). V.D.M. has filed a patent on the reverse genetic system and reporter SARS-CoV-2.

10.1126/science.abn3781

BIOCATALYSIS

Teaching natural enzymes new radical tricks

Cytochromes P450 were engineered to conduct abiotic, stereoselective radical reactions

By **Yang Zhang** *and* **Paweł Dydio**

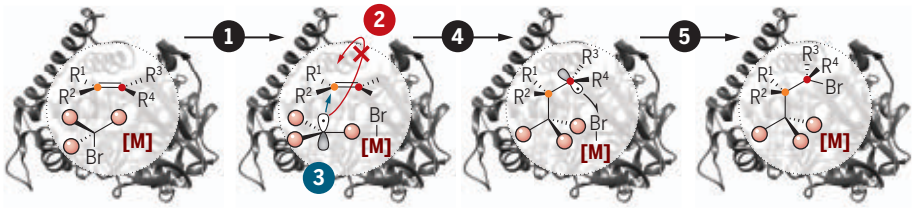
Chemical reactions that are difficult to execute with small-molecule catalysts can at times be executed selectively and efficiently with natural enzymes (*1*), reengineered enzymes (*2*), or even "artificial" enzymes (*3*). Organic reactions that occur through free-radical intermediates are particularly difficult to control with small-molecule catalysts, but their enzymatic counterparts are rare as well (*4*). On page 1612 of this issue, Zhou *et al.* (*5*) reported an elegant and powerful strategy to repurpose cytochrome P450, heme-based enzymes that natively conduct selective oxidation reactions, to catalyze an atom-transfer radical cyclization (ATRC), a free-radical reaction that thus far was only known to proceed unselectively in the presence of small-molecule catalysts. Directed evolution that iteratively accumulates constructive mutations to the protein scaffold (*2*, *3*) created variants of P450 enzymes that execute highly stereocontrolled ATRC. These results open a new forefront in unnatural radical biocatalysis (*4*).

Free-radical reactions that quickly build up molecular complexity, such as atom-transfer radical additions, are of fundamental importance in organic synthesis. However, the control of enantio- and diastereoselectivity of the formed products represents a long-standing challenge (*6*) and limits their utility in the synthesis of fine chemicals such as pharmaceuticals and agrochemicals. The intrinsic difficulty in controlling stereoselec-

From oxidation to radical reactions

Zhou *et al.* used directed evolution to repurpose P450 enzymes, which natively perform selective oxidations, so that they perform stereoselective atom-transfer radical cyclization reactions instead. Different evolved enzymes were created so that different stereoisomers could be produced.

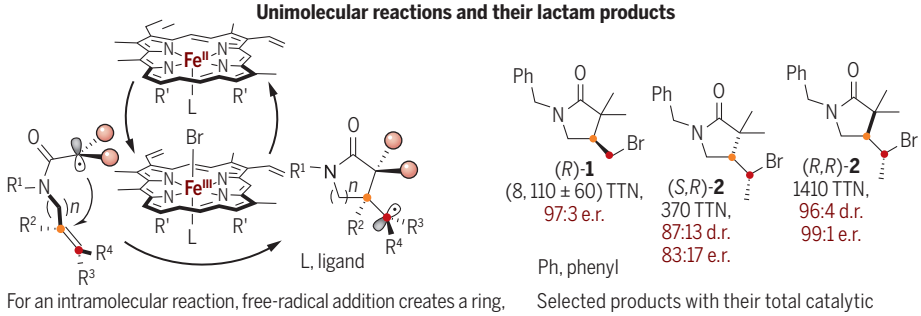
Enzymatic free-radical addition



1 Free-radical formation
2 Unfavored attack (→)
3 Favored attack (→)
4 Carbon-carbon bond formation
5 Carbon-bromine bond formation

A free radical forms when the heme metal center [M] abstracts Br from one substrate. The enzyme positions the double bond of the other substrate so that free-radical addition occurs from one side. Attack from the other side (shown as X) is prevented so that the process is stereoselective; R's are various substituents.

Unimolecular reactions and their lactam products



For an intramolecular reaction, free-radical addition creates a ring, and this intermediate cyclization step is enabled by a bromine shuttling mediated by an Fe redox center of the heme cofactor.

Selected products with their total catalytic turnover numbers (TTNs) and enantiomeric (e.r.) and diastereomeric (d.r.) ratios are shown.

GRAPHIC: KELLIE HOLOSKI/*SCIENCE* BASED ON Y. ZHANG AND P. DYDIO

tivity relates to the mechanism of these reactions. Upon the catalyst-mediated formation of free radicals, these key intermediates can diffuse out of the sphere of the catalysts before they can impose selectivity on subsequent stereoselectivity-determining steps. Some stereo-discriminative reactions have been reported that used noncovalent interactions between tailored catalysts and specially selected radical intermediates, but these strategies remain typically inapplicable to common unbiased substrates (*6, 7*).

Zhou *et al.* hypothesized that executing free-radical chemical transformations within an active site of an enzyme could enable controlling stereoselectivity by the intimate interactions between the reacting intermediates and the substrate binding site of the protein scaffold (see the figure). This strategy mimics the excellent stereocontrol of natural radical enzyme reactions (*8*). They used cytochromes P450 that contain the heme cofactor previously shown to mediate other atom-transfer radical reactions (*9*). These enzymes have been engineered to accommodate several nonnatural (albeit nonradical) reactions with high stereoselectivity and excellent activity (*2, 3, 10, 11*).

Initial evaluation of a range of heme and nonheme iron-dependent enzymes identified promising activity of previously reported P411-CIS T438S (*2*), a variant of P450 BM3 bearing 15 mutations, which furnished the ATRC γ-lactam product **1** with 60:40 enantiomeric ratio (e.r.). This modest stereoselectivity validated the initial hypothesis, and five rounds of site-saturated mutagenesis targeting the residues near the active site delivered a highly stereoselective and active variant $P450_{ATRCase1}$ bearing in total 20 mutations. $P450_{ATRCase1}$ formed the target ATRC product (*R*)-**1** with excellent selectivity of 97:3 e.r. and high productivity of >8000 total turnovers.

Formation of only one enantiomer of a product is often a drawback of enzymatic synthesis. Zhou *et al.* also synthesized (*S*)-**1**, the mirror image of (*R*)-**1**, selectively in the presence of $P450_{ATRCase2}$, an independently evolved variant of P450 BM3 bearing 23 mutations to the protein scaffold. Most notably, the engineered enzymes can accommodate other substrates. Not only was a range of γ-lactam products bearing different substituents or other aromatic rings in place of the phenyl ring formed, but also β- and δ-lactam products were produced stereoselectively. As a result, $P450_{ATRCases}$ provide access to these important structural motifs present in many bioactive molecules.

University of Strasbourg, CNRS, ISIS UMR 7006, 8 allée Gaspard Monge, 67000 Strasbourg, France.
Email: dydio@unistra.fr

For starting materials with an internal double bond bearing different substituents, two new stereocenters are created in the reaction. Both the initial addition of the free radical to the double bond (that is, the C-C bond-forming radical addition step) and the subsequent C-Br bond-forming halogen rebound step, are stereodetermining. Thus, the product could be formed as four different stereoisomers. Zhou *et al.* showed that both steps can be precisely controlled within the active site of the evolved variants of P450 BM3. Either diastereomer (that is, non–mirror image stereoisomer) of the product, such as (*S*,*R*)-**2** or (*R*,*R*)-**2** (see the figure), can be formed with high enantio- and diastereoselectivity. This extent of reaction stereocontrol also typically inaccessible with small-molecule catalysts (*6*).

These reactions proceeded efficiently in whole cells containing these enzymes without any additional manipulation or further purification. Whole-cell catalysts facilitate directed evolution and enable rapid evaluation of hundreds of variants should the stereoselectivity of the reaction need to be further improved for any substrate. The activity of $P450_{ATRCases}$ in a whole cell also opens possibilities for devising unnatural biosynthetic pathways that directly combine natural and unnatural enzymatic transformations to yield unnatural products (*12*).

The repurposing of metalloenzymes to control the reactivity of cumbersome radical intermediates reported by Zhou *et al.* suggests new opportunities for controlling other types of the atom-transfer radical addition reactions. Intermolecular transformations are arguably the most challenging but synthetically appealing. Given the breadth of free-radical reactions, these findings set the stage for the development of a myriad of stereoselective processes. ■

REFERENCES AND NOTES

1. G. A. Strohmeier, H. Pichler, O. May, M. Gruber-Khadjawi, *Chem. Rev.* **111**, 4141 (2011).
2. Y. Yang, F. H. Arnold, *Acc. Chem. Res.* **54**, 1209 (2021).
3. F. Schwizer *et al.*, *Chem. Rev.* **118**, 142 (2018).
4. T. K. Hyster, *Synlett* **31**, 248 (2020).
5. Q. Zhou, M. Chin, Y. Fu, P. Liu, Y. Yang, *Science* **374**, 1612 (2021).
6. M. P. Sibi, S. Manyem, J. Zimmerman, *Chem. Rev.* **103**, 3263 (2003).
7. R. S. J. Proctor, A. C. Colgan, R. J. Phipps, *Nat. Chem.* **12**, 990 (2020).
8. C. M. Jäger, A. K. Croft, *ChemBioEng Rev* **5**, 143 (2018).
9. K. J. Rodriguez *et al.*, *ACS Macro Lett.* **7**, 1111 (2018).
10. X. Ren, R. Fasan, *Curr. Opin. Green Sustain. Chem.* **31**, 100494 (2021).
11. S. N. Natoli, J. F. Hartwig, *Acc. Chem. Res.* **52**, 326 (2019).
12. J. Huang *et al.*, *Nat. Chem.* **13**, 1186 (2021).

ACKNOWLEDGMENTS

We acknowledge the European Research Council (ERC Starting Grant no. 804106), the French National Research Agency (ANR IdEx, and ANR LabEx "Chemistry of Complex Systems"), and the Frontier Research in Chemistry Foundation.

10.1126/science.abm8321

IMMUNOLOGY

Fibrin sparks inflammation in the oral mucosa

Fibrin deposits in the oral mucosa trigger neutrophil activation and bone destruction

By **Tommaso Vicanolo** *and* **Andrés Hidalgo**

Inflammation of the oral mucosa afflicts about half a billion individuals worldwide and is responsible for the destruction of the periodontium (the soft tissue and alveolar bone around the teeth). Periodontitis, a severe form of this disease, has been associated with dysbiosis of the oral microbiota (*1*) and with local deposition of fibrin, a fibrous protein involved in blood clotting. Periodontitis is also associated with loss-of-function mutations in the gene encoding plasminogen (PLG), a protease that degrades fibrin and prevents thrombosis. Mechanistic links between microbial and clotting elements and inflammation of the oral mucosa have remained enigmatic given the traditional view that clotting factors function inside blood vessels. On page 1575 of this issue, Silva *et al.* (*2*) show that deposits of fibrin appear in the oral mucosa in a microbiota-dependent manner and function as obligatory substrates for full activation of neutrophils—the most aggressive and tissue-destructive type of leukocyte.

Fibrinogen circulates in the blood in large amounts released by the liver, but it is also produced by other cells, including the mucosal epithelium (*3*). Because fibrin, the cleavage product of fibrinogen, associates both with thrombosis and with inflammation, controlling fibrin concentrations is important to preserving organismal homeostasis. The elimination of fibrin deposits is mediated by the protease plasmin, itself derived from proteolytic cleavage of its precursor protein PLG. In addition to periodontitis, defective fibrinolysis caused by PLG deficiency has been associated with ocular disease (conjunctivitis) as well

Area of Cell and Developmental Biology, Centro Nacional de Investigaciones Cardiovasculares, Madrid, Spain.
Email: ahidalgo@cnic.es

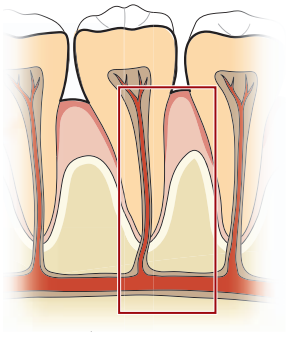
as pulmonary, vaginal, and gastrointestinal inflammation (*4*, *5*), which suggests that fibrin can accumulate in multiple organs and damage mucosal tissue if not properly degraded.

Using mice expressing a catalytically inactive form of PLG, mice lacking the enzymes responsible for processing PLG to its active form, and mice deficient in both *Plg* and *Fga* (the gene encoding one of the fibrinogen chains), Silva *et al.* elegantly demonstrate that PLG prevents oral inflammation by specifically preventing the accumulation of fibrin. But how does a fibrillary protein involved in clotting trigger inflammation? Silva *et al.* show that fibrin deposits are indispensable for neutrophil activation in the oral mucosa. Using a combination of genetic models and in vitro assays, the authors demonstrate that direct engagement of fibrin by the Mac-1 receptor expressed on neutrophils induces the production of reactive oxygen species (ROS) and neutrophil extracellular traps (NETs, DNA-protein fibers), which cause tissue damage by chemical, lytic, or enzymatic activity. Mac-1 is expressed by all myeloid cells, and therefore it cannot be ruled out that fibrin deposits also convert monocytes and macrophages into effectors of mucosal inflammation. An important unsolved question is the identity of the cells that locally produce fibrin deposits in the mucosa, because anomalous activity from these cells—for example, that induced by microbial dysbiosis—causes excessive fibrin production and inflammation (see the figure).

An exciting implication from the study of Silva *et al.* is that neutrophils recruited to the site of inflammation are not by default activated and primed for tissue destruction. This challenges textbook neutrophil biology, which places recruitment as the main, if not the exclusive, bottleneck for overt inflammation (*6*). These findings suggest that local cues strategically deposited in the tissue are needed for full-blown activation and release of the toxic arsenal of these cells. Because the activity of neutrophils is generally protective against microorganisms that occasionally invade tissues, the study also implies that deposits of fibrin must be constantly produced and degraded at the interface of the tissue and adjacent commensal microbiota of healthy individuals. Examination of the formation and degradation dynamics of such proinflammatory fibrin hubs across tissues may unearth mechanisms that keep inflammation in check. Neutrophils can be detected in most naïve tissues, including the oral cavity (*4*, *5*), where they acquire distinct properties (*7*). It is therefore conceivable that not only the formation of these deposits, but also the presence of neutrophil populations with exaggerated responses to fibrin, underlie pathological inflammation of the mucosa.

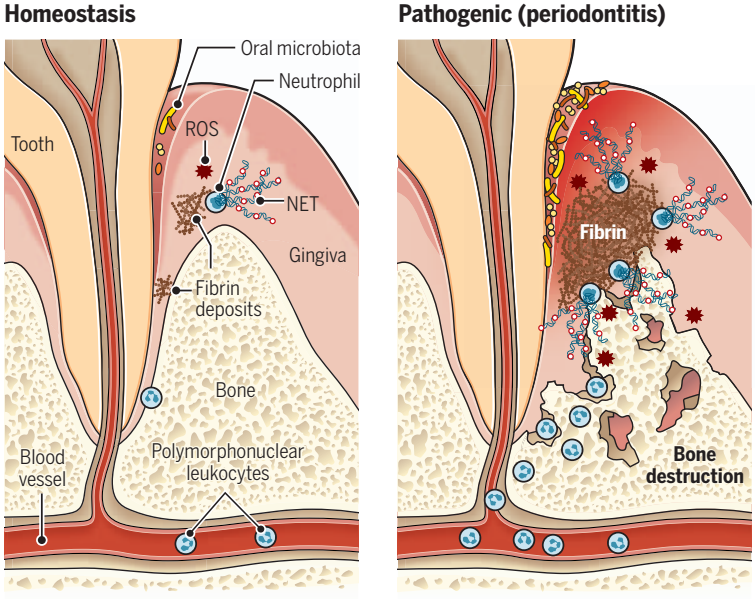
Recent studies have shown that fibrin is a key structural element of innate immunity in serosal cavities in the body, where it participates in pathogen removal and tissue repair by trapping bacteria together with immune cells in fibrillar aggregates (*5*, *8*, *9*). These studies substantiate the mechanistic (and possibly evolutionary) connection between coagulation and the innate immune system, a concept that has been reinforced by the discovery of macrophages with platelet-like behavior in large body cavities (*5*). The work of Silva *et al.* extends this fascinating biology by showing that fibrin functions as a physical hub to launch neutrophil responses in the oral mucosa at the right place and time and by demonstrating that resolving inflammation demands the removal of these proinflammatory substrates.

These findings raise additional questions regarding the actual mechanisms of injury. For example, NETs are well-established drivers of vascular disease through either enzymatic or cytolytic activity (*10*). However, the mechanisms by which these DNA scaffolds cause bone erosion are unclear: They could mediate direct degradation of the bone or indirectly activate osteoclasts (bone-degrading cells). This could occur by the degradation of proteins that prevent bone loss (*11*) by the proteases that decorate the NET. Direct evidence that immune-regulatory fibrin substrates form in other mucosal surfaces (such as the gut, skin, or vagina), or the possible existence of analogous hubs involving different structural proteins in other tissues, warrants further research. Developing therapies that exploit the fibrin-neutrophil axis may provide much-needed relief to those living with mucosal inflammation. ■



Inflammation of the oral mucosa

Under homeostatic conditions, microbiota may induce the deposition of fibrin. This stimulates the activation of neutrophils, which release toxic reactive oxygen species (ROS) and neutrophil extracellular traps (NETs). Degradation of fibrin deposits by plasmin (from plasminogen) is required to prevent excessive immune activation, chronic inflammation, and bone destruction, which occur during periodontitis.



GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

REFERENCES AND NOTES

1. G. Hajishengallis, *Nat. Rev. Immunol.* **15**, 30 (2015).
2. L. M. Silva *et al.*, *Science* **374**, 1575 (2021).
3. D. W. Williams *et al.*, *Cell* **184**, 4090 (2021).
4. J. W. Chadwick *et al.*, *J. Leukoc. Biol.* **110**, 187 (2021).
5. J. Zindel *et al.*, *Science* **371**, eabe0595 (2021).
6. S. Nourshargh, R. Alon, *Immunity* **41**, 694 (2014).
7. I. Ballesteros *et al.*, *Cell* **183**, 1282 (2020).
8. A. Vega-Pérez *et al.*, *Immunity* **54**, 2578 (2021).
9. N. Zhang *et al.*, *J. Exp. Med.* **216**, 1291 (2019).
10. C. Silvestre-Roig *et al.*, *Nature* **569**, 236 (2019).
11. J. Shin *et al.*, *Sci. Transl. Med.* **7**, 307ra155 (2015).

ACKNOWLEDGMENTS

A.H. is a paid consultant for Flagship Pioneering. T.V. is supported by a fellowship from the La Caixa Foundation (ID100010434). A.H. is supported by RTI2018-095497-B-I00 from Ministerio de Ciencia e Innovación (MICINN), HR17_00527 from Fundación La Caixa, the Transatlantic Network of Excellence (TNE-18CVD04) from the Leducq Foundation, and FET-OPEN (no. 861878) from the European Commission.

10.1126/science.abn0399

COVID-19 vaccine breakthrough infections

Vaccine efficacy wanes over time but can be fully restored with a booster dose

By **Ravindra K. Gupta**[1,2,3] *and* **Eric J. Topol**[4]

The pivotal phase 3 clinical trials of the two-dose messenger RNA (mRNA) vaccines, among the largest ever conducted, led to the notable finding of ~95% efficacy for prevention of symptomatic COVID-19 2 months after the second dose. Adenovirus vectored vaccines showed lower protection against infection with severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) but achieved >90% protection against severe disease. No vaccines protect against all infections, and very few achieve such a high level of protection as that of the COVID-19 vaccines. In the early months after vaccinations at scale began, around January 2021, post-vaccination infections (or breakthrough infections) were rare, accounting for <1% of COVID-19 cases, and only ~0.1% resulted in hospitalization or death in high-income countries (*1*, *2*). However, by 5 to 6 months after vaccinations began, this pattern changed. Even before vaccine introduction, it was anticipated that a third, booster dose would be necessary to preserve efficacy, but when that would be needed was uncertain.

Variants of SARS-CoV-2 with multiple immune-escape and infectivity-enhancing mutations (particularly in the spike protein, which facilitates infection of human cells) likely arise after chronic infection (*3*). Some, such as the Alpha variant, showed increased infectivity and transmissibility, whereas other variants, such as Beta, were less sensitive to neutralization by vaccine- and infection-induced antibodies. As Delta became prevalent in Israel, the United Kingdom, Qatar, and the United States, there were multiple reports of a substantial increase in breakthrough infections after mRNA and adenovirus vectored vaccination. The Israel Ministry of Health reported vaccine effectiveness of 40% against symptomatic infections 4 to 6 months after the second dose, representing a substantial decline (see the figure). Although initially it was unclear whether this was due to waning immunity over time or the more transmissible Delta variant, it became apparent that time itself was a key driver, with attrition of efficacy seen in the participants of initial clinical trials (*4*). Waning immunity occurred, to a variable extent, after all vaccines studied to date, and loss of protection was likely amplified by increased prevalence of Delta.

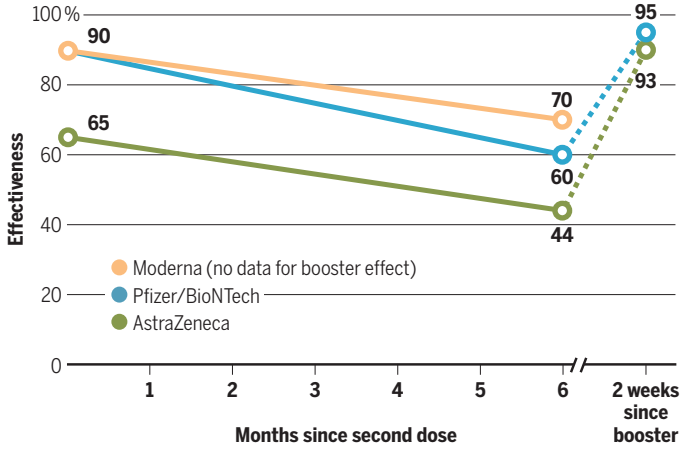
The Delta variant contains mutations in the spike protein that are divergent from the three prior variants of concern, Alpha (B.1.1.7), Beta (B.1.351), and Gamma (P.1). SARS-CoV-2 is an enveloped virus, taking lipid bilayer from infected cells. The infected cell produces the viral spike protein after translation from the viral RNA template, and spike is embedded in the lipid bilayer surrounding the virus core. The spike protein has two main regions, S1 and S2. During transport of spike to the plasma membrane, it is cleaved by cellular furin proteases at the furin cleavage site (FCS) between the S1 and S2 regions. The two cleaved pieces then associate with each other to form semimature spike at the cell surface. For optimal cellular infection of the produced virions, S1S2 also needs processing at the S2′ site adjacent to the FCS; this processing is carried out by transmembrane protease serine 2 (TMPRSS2) in the plasma membrane. Spike is then able to interact with the host cell receptor, angiotensin-converting enzyme 2 (ACE2), through the receptor binding domain (RBD) (in S1) to drive efficient fusion of the viral membrane with the host cell membrane for entry of the virus.

One mutation, P681R (Pro^{681} → Arg), is located in the FCS and is specific to Delta. The mutation is associated with increased cleavage of spike into S1 and S2 fragments (*5*). The role of P681R might be related to other spike mutations—for example, in the amino-terminal domain (NTD) of S1, where Delta bears a deletion of amino acids 157–158—as well as T19R, G142D, and R156G mutations. These NTD mutations lead to substantial rearrangement of the NTD that may have allosteric effects on the RBD and/or promote binding with additional cellular receptors to increase infectivity. Spike mutations that increase infectivity could enable virus to rapidly attach and infect epithelial respiratory cells, avoiding the relatively sparse neutralizing antibodies in the mucosa. Furthermore, Delta's spike protein can achieve membrane fusion far more efficiently than can other variants (*5*). This ability to fuse cells to generate syncitia (multinucleated cells) might enable virus to propagate from one cell to another without needing to exit the cell, avoiding exposure to neutralizing antibodies.

Delta has also demonstrated moderate evasion from neutralizing antibodies, which appears to be partly related to the RBD mutation L452R with a less clear contribution from T478R in the RBD. Mutations in the Delta NTD have also been shown to reduce recognition by NTD-specific neutralizing antibodies. Together, immune evasion and increased replication likely underpin Delta's ability to cause reinfection and vaccine breakthrough.

Vaccine effectiveness over time

Two doses of messenger RNA (mRNA) or adenovirus vectored COVID-19 vaccines elicit high levels of protection from symptomatic disease, but this wanes over time. Emerging studies show that a third dose (booster) of the same type can restore effectiveness to >90%. Data are averages for Delta variant from multiple studies.



GRAPHIC: N. DESAI/*SCIENCE*

[1]Cambridge Institute of Therapeutic Immunology and Infectious Disease (CITIID), Cambridge, UK. [2]Department of Medicine, University of Cambridge, Cambridge, UK. [3]Africa Health Research Institute, KwaZulu-Natal, South Africa. [4]Scripps Research, La Jolla, California, USA. Email: rkg20@cam.ac.uk; etopol@scripps.edu

A consistent finding across studies has been the high viral load associated with Delta infections, no matter whether they occur among unvaccinated or vaccinated individuals. Recent transmission studies of the Delta variant have also revealed some distinct features, particularly a faster onset of illness and clearance. Faster clearance of the virus and a shorter duration of infectivity was noted in a study of vaccinated compared with unvaccinated people (*6*). The magnitude of transmissibility by individuals with Delta breakthrough infections appears to be approximately half that compared with unvaccinated individuals (*7*), which is supported by reduced culture-positive virus in some vaccinated individuals with high viral loads [as assessed by <25 cycle threshold (Ct), the number of cycles required for a positive result in real-time polymerase chain reaction (RT-PCR) tests]. A recent report of nearly 140,000 people who were contacts of individuals with RT-PCR–confirmed COVID-19 showed that both the AstraZeneca and Pfizer/BioNTech vaccines suppressed transmission, but their capacity to do so was markedly lower for Delta compared with Alpha and lower in people vaccinated with AstraZeneca compared with the Pfizer/BioNTech vaccine (*8*).

Furthermore, two independent reports have confirmed that high amounts of viral RNA (Ct <25) occur in asymptomatic Delta breakthrough infections, and that these individuals could be transmitting SARS-CoV-2 to others (*9*). However, distinction should be made between infectious virus and Ct value, and the relationship between the two in vaccinated versus unvaccinated individuals needs further evaluation. Overall, transmission from vaccinated individuals is increased by the Delta variant, compared with previous strains, according to the setting and length of time elapsed from initial vaccination.

Time appears to be the key driver of the post-vaccination reduction in effectiveness, as demonstrated from a study of 3.4 million members of the Kaiser Permanente health care organization that found a similar pattern of decline in immunity against multiple variants from 2 months after the second dose (*4*). Although many studies have confirmed a reduction in serum concentrations of neutralizing antibodies from 4 to 6 weeks after vaccination (*10*, *11*), the picture is less clear for $CD4^+$ and $CD8^+$ T cell responses, with studies showing small changes consistent with the development of immune memory. The clinical waning of immunity after the first 2 months is particularly notable in people over 60 years of age, in whom susceptibility increased for both symptomatic infections and hospitalizations, as first noted in Israel and later confirmed in multiple US Centers for Disease Control and Prevention (CDC) reports. Circulating and tissue-neutralizing antibodies are expected to wane in a few months despite maintenance of specific memory B cell populations in the circulation. T cells, normally mobilized in response to infection, are thought to protect individuals from severe disease. That hospitalizations were increasingly noted in people of advanced age with breakthrough infections is consistent with poorer B cell and T cell responses to vaccination in older people, as shown in a study examining responses to the Pfizer/BioNTech mRNA vaccine (*12*). More studies exploring the trajectory of vaccine-induced cellular responses over time and according to age are needed.

Although the clinical trials of mRNA vaccines used a short time interval between two doses, 3 to 4 weeks for Pfizer/BioNTech and Moderna, shortage of the vaccines in many countries led to adoption of 8- to 16-week spacing. Scotland and Canada found that extended spacing of mRNA vaccines led to >80% effectiveness against symptomatic infection in the first few weeks after vaccination. Moreover, the most substantial drop-off in vaccine effectiveness (before the Delta variant became dominant) was observed using a 3- to 4-week dosing interval, such as in Israel, the United States, and Qatar. A direct comparison of short and long dose spacing for the Pfizer/BioNTech vaccine demonstrated that a 16-week spacing between doses resulted in optimal humoral immune responses (*13*). Administration of two mRNA vaccine doses, closely spaced by 3 to 4 weeks, may have acted as a primary immunization—maximally inducing neutralizing antibodies but compromising durable immunity. That compromise may take the form of both humoral and cellular immunity waning in high-risk individuals, such as the elderly or immunocompromised as early as 2 months after the second dose.

Immunologic studies of responses to boosters, given 6 months after the last vaccine dose, have uniformly shown the induction of very high amounts of neutralizing antibodies, which correlates with protection from breakthrough infection. In Israel, where more than 1.1 million people over 60 years of age received an mRNA vaccine booster dose 6 months after the second dose, restoration of more than 90% effectiveness against severe COVID-19 was achieved (*14*). The restoration of vaccine effectiveness against hospitalizations and deaths with a booster dose was subsequently demonstrated for adults aged 40 years and older. A large, placebo-controlled randomized trial of the Pfizer/BioNTech booster indicated 95% efficacy, with reduction of symptomatic infections across all adults, age 18 and over (*15*). Data are lacking for other vaccines and the durability of this effect. With continued circulating virus over time, it is likely that improved efficacy of a booster dose will be further demonstrated, in addition to reduced transmission and fewer cases of Long Covid (which can probably occur after vaccine breakthrough infection).

The high transmission rates observed in North America and Europe, where vaccine coverage is greatest, portends selection of vaccine escape variants of SARS-CoV-2 that could overcome some of the protection against severe disease. These variants are likely to arise during chronic infections in those with suboptimal vaccine responses, such as people who are immune-compromised, or where vaccine waning has occurred. New variants may evolve from Delta or may be radically different and could even be recombinants of variants due to mixed infections within individual hosts. Recent identification of B.1.1.529 (Omicron) with multiple spike mutations in southern Africa is a reminder of the ongoing threat posed by SARS-CoV-2. Continued transmission in highly vaccinated populations underscores the need for expansion of vaccination across age groups while maintaining nonpharmacological interventions, such as mask wearing. Investigation of intranasal vaccine preparations as a means of preventing breakthrough infection, development of pan-sarbecovirus vaccines, and exploration of the potential for antiviral medications should also be explored to limit transmission. ■

REFERENCES AND NOTES

1. E. J. Haas *et al.*, *Lancet* **397**, 1819 (2021).
2. F. P. Polack *et al.*, *N. Engl. J. Med.* **383**, 2603 (2020).
3. S. A. Kemp *et al.*, *Nature* **592**, 277 (2021).
4. S. Y. Tartof *et al.*, *Lancet* **398**, 1407 (2021).
5. P. Mlcochova *et al.*, *Nature* **599**, 114 (2021).
6. P. Y. Chia *et al.*, *medRxiv* 10.1101/2021.07.28.21261295 (2021).
7. A. Singanayagam *et al.*, *Lancet Infect. Dis.* (2021).
8. D. W. Eyre *et al.*, *medRxiv* 10.1101/2021.09.28.21264260 (2021).
9. C. B. Acharya *et al.*, *medRxiv* 10.1101/2021.09.28.21264262 (2021).
10. M. J. van Gils *et al.*, *medRxiv* 10.1101/2021.09.27.21264163 (2021).
11. Y. Goldberg *et al.*, *N. Engl. J. Med.* NEJMoa2114228 (2021).
12. D. A. Collier *et al.*, *Nature* **596**, 417 (2021).
13. A. Tauzin *et al.*, *medRxiv* 10.1101/2021.09.17.21263532 (2021).
14. Y. M. Bar-On *et al.*, *N. Engl. J. Med.* **385**, 1393 (2021).
15. CDC, "Efficacy & Safety of BNT162b2 booster—C4591031 2 month interim analysis" (2021); www.cdc.gov/vaccines/acip/meetings/downloads/slides-2021-11-19/02-COVID-Perez-508.pdf.

ACKNOWLEDGMENTS

R.K.G. is supported by Wellcome WT108082AIA and has received honoraria for educational activities from Janssen and GlaxoSmithKline. E.J.T. is supported by the National Institutes of Health (UL1TR002550).

10.1126/science.abl8487



POLICY FORUM

BIOTECHNOLOGY

Enabling commercial success of industrial biotechnology

Commercial-scale research translation has been muted

By **Bradley W. Biggs[1], Hal S. Alper[2], Brian F. Pfleger[3], Keith E. J. Tyo[1], Christine N. S. Santos[4], Parayil Kumaran Ajikumar[4], Gregory Stephanopoulos[5]**

The industrial biotechnology sector—commercial-scale manufacturing of chemical products by use of cellular or molecular biocatalysts—lies at a crossroads. Metabolic engineering aided by tools of systems and synthetic biology has expanded the scope of possible chemical products that can be made with this approach and the avenues available for improvements in strain engineering. Academic interest is strong and growing, with several high-profile demonstrations of metabolic engineering (*1*, *2*). However, few academic successes have become commercial successes. The low success rate of translating research stifles job growth, prevents new products from reaching consumers, and leaves societal and ecological benefits of industrial biotechnology unrealized. We highlight three key areas that need greater attention, recognition, and support: (i) policy to promote translational science and prioritize sustainability, (ii) investment in technology and infrastructure to enable translational research and education, and (iii) prioritization of biomanufacturing workforce training and education.

[1]Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL, USA. [2]McKetta Department of Chemical Engineering, University of Texas at Austin, Austin, TX, USA. [3]Department of Chemical and Biological Engineering, University of Wisconsin–Madison, Madison, WI, USA. [4]Manus Bio, Cambridge, MA 02138, USA. [5]Department of Chemical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02142, USA. Email: pkaji@manusbio.com; gregstep@mit.edu

Industrial biotechnology is especially suited for applications with favorable sustainability metrics because of the ability of biological systems to handle a variety of renewable feedstocks, including complex and heterogeneous feedstocks such as nonfood plant biomass (*3*) and single-carbon gasses (*4*). Because the specificity of enzymes can enable one-pot synthesis of complex chemicals (*5*), biological processes can produce substantially less waste and require less energy when compared with multistep chemical synthesis (*6*). Together, these advantages—if successfully leveraged—could provide security for material goods supply, environmental benefits in chemical product synthesis, and even the potential for the synthesis of new chemicals with broad new uses as materials and medicines.

POLICY

Industrial biotechnology competes with chemical processing, a multitrillion-dollar industry with strong patent protection, a well-trained workforce, economies of scale, and depreciated facilities. Environmental sustainability, a main advantage of industrial biotechnology, is commonly undervalued by society and regulators. If valuation does not incorporate the ecological impact of the approach to chemical or biochemical synthesis, industrial biotechnology will be relegated to only the markets and products that cannot be feasibly obtained any other way. High-volume, low-margin products such as polymers, industrial chemicals, and fuels will continue to be produced through traditional chemical processing, forfeiting the sustainability benefits of industrial biotechnology.

These 300-liter tanks can be used to mix and sterilize media prior to fermentation, or to mix buffers and hold material generated in downstream processing, at the Advanced Biofuels and Bioproducts Process Development Unit, part of Lawrence Berkeley National Laboratory.

One solution would be to require any synthesis process (chemical or biological) to restore the environment to its original state, from before the synthesis process, including gases used and generated. Requiring equal life-cycle impact for the two approaches could help balance competition between chemical and biological processes for high-volume chemicals while improving environmental conditions regardless of which approach is ultimately used. Currently, more than 40 countries account for greenhouse gas emissions, as determined from life cycle analysis, to ensure that new, sustainable commercialization efforts are on a level playing field with mature, higher-emission businesses (*7*).

Emerging technologies without an established pipeline to commercialization can have a difficult time acquiring funding for initial technological translation of academic discoveries because existing funding sources can either have too low of a risk tolerance or misaligned funding priorities. This is the case for synthetic biology and metabolic engineering, with the science progressing much faster than is the infrastructure for translating fundamental discovery. This type of early-stage research translation has special requirements, including funding through small, targeted grants to surgically tackle key initial questions and funding of multidisciplinary solutions that aim to reduce risk in the integrated system and solve complex problems. For example, many industrial biotechnology processes are built on a handful of microbial strains (chassis). Understanding, at a mechanistic level, the cellular response to the difference in stress conditions between the laboratory and scale would be broadly useful to the field.

This can be accomplished with existing funding structures when properly modified to address various shortcomings. For example, in the United States current Small Business Innovation Research and Small Business Technology Transfer research grant structures make formation of a company compulsory, which does not allow for exploration before company formation of the industrial potential of a technology. Although this mandate is satisfied by various companies, we can envision many technologies that are not ready for company formation yet show great commercialization potential if some outstanding risks were removed. Once risk is adequately reduced, such technologies can move further toward commercialization

PHOTO: BERKELEY LAB/ROY KALTSCHMIDT

through the formation of new companies or licensing to existing companies.

Related to this is the potential to relax restrictions on the involvement of principal investigators from universities and national laboratories in early stages of technology transfer. Although conflict-of-interest considerations exist, they could be mitigated with appropriate oversight and coordination among universities and funding agencies. Policies vary among institutions but should be reevaluated if found to create barriers to translational efforts.

Other adjustments to existing funding structures could help support the translation of emerging fields. There is a current focus on large multi-institutional, interdisciplinary center grants. In these funding vehicles, disparate expertise is gathered to address a complex problem. The basic premise for justifying these center grants is that by supporting a single center, they provide conditions in which solutions to large interdisciplinary problems can be sought more efficiently. It is our experience, however, that this premise is often false. Priorities and interests can be misaligned, with a subset of stakeholders disinterested or poorly engaged. For some translational projects, intermediate-sized grants (~$0.75 million to $1 million annually for 5 years) could allow small interdisciplinary teams to function with more flexibility, higher accountability, and closer connection of principle investigators to the goals of the program.

Further, such funding modifications would increase the necessity of translational research and commercialization and thus help create a culture of support and recognition for innovative solutions to critical bottlenecks in the practical application of new discoveries.

TECHNOLOGY AND INFRASTRUCTURE

An industrial biotechnology process comprises a biocatalyst (microbe or enzyme), a bioprocess that produces the product, and an appropriate downstream purification facility. Consider the diversity of biocatalysts (cells or enzymes), process environments (aerobic, microaerobic, or anaerobic), modes of operation (batch, fed batch, or continuous), and types of substrate (glucose, glycerol, methane, or lignocellulosic biomass). Successful scale-up of a biomanufacturing process means providing a bioprocess environment (exemplified in the design and operation of bioprocess equipment) ideally suited to the specific needs of the biocatalyst. This is not a trivial process, and failure to do so may waste an otherwise excellent biocatalyst. Bioprocess development is not straightforward, able to rise to the requirements of any biocatalyst.

List of academic pilot-scale facilities in the United States

FACILITY	LOCATION	MAX CAPACITY (LITERS)
CSL Behring Fermentation Facility (Huck Institutes of the Life Sciences, Pennsylvania State University)	University Park, PA	100
Biological Process Development Facility (University of Nebraska-Lincoln)	Lincoln, NE	150
Massachusetts Biomanufacturing Center (University of Massachusetts-Lowell)	Lowell, MA	200
Biotechnology Scale-Up Facility (University of Maryland)	College Park, MD	250
Advanced Biofuels Process Development Unit (Lawrence Berkeley National Lab)	Emeryville, CA	300
Synthetic Biomanufacturing Facility (Utah State University)	North Logan, UT	400
Bioexpression and Fermentation Facility (University of Georgia)	Athens, GA	750
Cell and Cell Product Fermentation Facility (Rutgers University)	Piscataway, NJ	800
Center for Biocatalysis and Bioprocessing (University of Iowa)	Coralville, IA	1000
Fermentation Facility (Iowa State University)	Aimes, IA	1000
Integrated Bioprocessing Research Laboratory (University of Illinois Urbana-Champaign)	Urbana, IL	1500
Bioeconomy Institute (Michigan State University)	Lansing, MI	3800
Integrated Biorefinery Research Facility (National Renewable Energy Lab)	Golden, CO	9000
National Corn-to-Ethanol Research Center (Southern Illinois University-Edwardsville)	Edwardsville, IL	~22,000

With respect to required infrastructure, the greatest need is facilitated access to and use of pilot-scale fermentation facilities. Pilot-scale facilities are essential for the risk-reduction process, particularly for evaluation of biocatalyst performance in scale-up environments, which present challenges that are not found in laboratory-scale studies. Several products have met challenges in this process. Artemisinin production has faced scale-up challenges, with fermentation-based synthesis being consistently more expensive than extraction from farmed plants (*8*). Astaxanthin faced scale-up challenges after 2 years of production in a toll manufacturing plant (*9*). Upon attempting to scale up to a facility that was four times larger, poor batch performance required troubleshooting that substantially delayed production. The deficit of facilities and inability to reduce risk are major contributors to the "valley of death" between academic demonstration of a technology and its commercial-scale deployment. Reducing risk in scale-up studies is expensive and requires equipment that is not typically part of an academic setting. In addition, it is considered too applied for government support and too premature for industrial investment.

As a first step, we have compiled a list of pilot-scale (approximately 100- to 20,000-liter units) facilities at universities and national laboratories in the United States (see the table). Readily noticeable from this list is their limited capacity, both max capacity at any given site and overall capacity. However, this list can provide a starting point for a well-organized registry or system, such as the EPP Network (www.eppnetwork.com) and Pilots4U (https://biopilots4u.eu) in the European Union. A second consideration for these existing sites is their varied staffing, downstream processing capabilities, modes of operation, and user fees. With some exceptions, the primary functionality of these sites is a contract manufacturing model, in which the operations team is decoupled from and not invested in proving out a biocatalyst. It also provides minimal training and education. Others have more of a biopharma focus, such as those associated with the Advanced Mammalian Biomanufacturing Innovation Center, whose mission differs from that of industrial biotechnology. The solution will likely take a combination of greater pilot-

scale capacity, intentionally providing funding for maintaining necessary staffing, creation of a network to improve accessibility, and greater incorporation of these facilities into training programs.

Current academic training is primarily focused on biocatalyst development with minimal bioprocess optimization, owing in part to the lack of access to pilot-scale facilities. Along with restricting bioprocess development, training of students is similarly hampered. Without training in bioprocess and technological development, and the equipment needed to do so, biocatalytic advances that culminate in high-performing enzymes or microbes will continue to be very difficult to deploy commercially because the lack of suitable bioprocess infrastructure will prevent realization of laboratory-demonstrated TRY (titer, rate, and yield) at scale and therefore economies of scale.

A parallel need, although at the opposite end of the spectrum, exists for reliable scale-down methodologies and equipment. Modern genetic methods can generate a wealth of microbial variants in the quest for overproducing strains. The potential of these strain candidates must be assessed under conditions that resemble as closely as possible the actual bioprocess environment. This task requires small-scale equipment to evaluate multiple strains in parallel and thus increase throughput and reduce cost. Several miniaturized systems have been developed recently (*10*), but further validation is needed to ensure that they accurately recreate scale-up conditions. To accomplish this, access to scale-up facilities is required, and accurate strain characterization may require "omics" systems biology measurements in addition to macroscopic process measurements. These measurements will allow faithful recreation of the commercial-scale environment with the small-scale systems. One may never be able to create an accurate scale-down system if the large-scale system is undercharacterized. In particular, there is an acute need for greater emphasis on scaled-down systems recapitulating cell stresses (and thus resulting phenotypes) of large-scale cultivation, such as shear effects, gradients of pH, temperature, dissolved gasses, substrates and nutrients, and other limitations of mass transfer. Once achieved, accurate scale-down systems are invaluable for more rapid strain engineering because they will allow the generation of datasets for testing numerous strain (genetic) variants under different conditions. This approach is in contrast to most current datasets that screen many parameters or conditions, but for only a few strains. If successful, these scale-down systems could represent a major advance in addressing the "test" step of the strain optimization cycle.

EDUCATION

There has been a decline in teaching biochemical engineering fundamentals among many universities. Courses on cell growth and product formation, bioenergetics, enzymology of single enzymes and bioreaction networks, bioreactor design, and bioseparations have been exchanged for emerging fields such as nanotechnology, molecular biotechnology, biomaterials, and drug delivery. This is part of a general shift toward the molecular scale and pure sciences at the expense of engineering emphasis (*11*). Yet, biochemical engineering topics are central to understanding cell function and to the design and successful operation of bioprocess equipment of any scale. The primary drivers of this educational shift are research funding, scientific journal recognition, and the overall reward system that emphasizes molecular over bioprocess research and development, often placing a priority on novelty over problem solving and real-world impact. This is a serious issue that impairs the overall enterprise of industrial biotechnology because graduates lack experience and training to fill key positions, such as those involved in fermentation and scale up of downstream processes (*12*).

This issue needs to be addressed at multiple levels. At academic institutions, one possibility is through increased or renewed investment in master's-level courses and training that will cover the basics of industrial biotechnology, particularly training students versed in engineering fundamentals (kinetics, thermodynamics, and transport phenomena), bioprocess (vessel design, operation, and sterility), and bioseparations. In these programs, an emphasis should be placed on solving open-ended design problems in multidisciplinary teams, acknowledging that complex solutions require diverse expertise (for example, bioreactor design and automation requires biochemical, mechanical, and electrical engineering skills). For example, the Masters of Bioprocess Engineering in the Department of Chemistry at the University of California, Berkeley provides hands-on experience through industrial internship and their partnership with the Advanced Biofuels Processing Demonstration Unit.

Training in these areas must be differentiated from that of pharmaceuticals because, for example, it is a different challenge to concentrate (remove water from) a protein solution than it is to concentrate chemical products. There also needs to be greater partnership in the formulation of these courses, across academia and industry. Very few engineering departments carry the full breadth of expertise to teach all necessary courses, but in the era of virtual classrooms, academic consortia could be formed to bring the necessary skills and provide broad coverage of the curriculum. With respect to industrial involvement, there would be benefit from increased opportunity for site visits, with co-op opportunities for hands-on experience, and the possible creation of modified or redacted case studies provided by industry for academic settings. And there would be benefit from investment at the community college level, helping train students for roles such as equipment operator.

With respect to professional recognition, translational research ought to be given a higher profile. For example, advancing titers of milligram to multigram per liter values at pilot plant scales or works that provide innovative downstream processing techniques for separation of challenging products involve ingenuity and novelty that often go unrecognized by academic reward systems. The reward system should be reevaluated to recognize the importance of problem-solving, as opposed to the present focus on novelty and curiosity, and this should be reflected in journal publications as well as research funding decisions. Novelty is important, but so are solutions to real problems that impair innovation and industrial productivity. Just like the National Science Foundation investment in the Massachussetts Institute of Technology Biotechnology Process Engineering Center undergirded the rise of the cell-culture industry, a similar program focused on industrial biotechnology could revitalize this critical industry today.

Ultimately, to reap the potential benefits of industrial biotechnology, appreciation must be given to, and investments made in, the innovation and ingenuity required for translating to practice emerging scientific breakthroughs. ■

REFERENCES AND NOTES

1. C. J. Paddon *et al.*, *Nature* **496**, 528 (2013).
2. P. K. Ajikumar *et al.*, *Science* **330**, 70 (2010).
3. R. Kumar, S. Singh, O. V. Singh, *J. Ind. Microbiol. Biotechnol.* **35**, 377 (2008).
4. F. Liew *et al.*, *Front. Microbiol.* **7**, 694 (2016).
5. N. J. Turner, *Nat. Chem. Biol.* **5**, 567 (2009).
6. J. D. Keasling, *Science* **330**, 1355 (2010).
7. World Bank Group, *State and Trends of Carbon Pricing 2019* (World Bank, 2019).
8. M. Peplow *C&EN Glob., Enterp* **96**, 29 (2018).
9. J. Lievense, "Scaling up industrial biotechnology," 7th Life Science Symposium (2016).
10. W. Olughu, G. Deepika, C. Hewitt, C. Rielly, *J. Chem. Technol. Biotechnol.* **94**, 647 (2019).
11. A. Varma, I. E. Grossmann, *AIChE J.* **60**, 3692 (2014).
12. J. Philp, *N. Biotechnol.* **40**, 11 (2018).

ACKNOWLEDGMENTS

This manuscript is the product of discussions at a workshop convened in Augusta, Georgia, USA, by Manus Bio in November 2019. For a resource-related workshop, the following website has been prepared: www.rcm-biomanufacturing.org. We gratefully acknowledge T. Papoutsakis, J. Liao, and C. Cooney for their careful reading of and feedback on this manuscript. K.E.J.T., C.N.S.S., P.K.A., and G.S. have financial interest in Manus Bio.

10.1126/science.abj5040



Paleontologist Sanaa Al-Sayed probes the remains of a *Mansourasaurus shahinae* in 2018.

PHOTO: REUTERS/MOHAMED ABD EL GHANY

PALEONTOLOGY

Revisiting paleontology's greatest hits

An otherwise solid dinosaur primer does little to convey the changing face of the field

By **Riley Black**

Dinosaurs garner esteem that is often reflected onto the people who search for, excavate, and study them, and therein lies a fundamental problem with the ever-increasing number of popular tomes about the "terrible lizards" hitting bookshelves. Even as the field of vertebrate paleontology pushes to become more inclusive, personages from decades past remain the only experts many members of the public encounter. Although there is a trove of dinosaurian information to recommend paleontologist Darren Naish's short encyclopedia *Dinopedia,* it does little to correct this antiquated view of who is, or can be, a paleontologist.

"For a group of animals that died out 65.5 million years ago, dinosaurs…have a remarkable grip on our imagination," Naish writes at the top of the book. What follows is a friendly and breezy tour of dinosaurs and what paleontologists have come to know about them.



Dinopedia
Darren Naish
Princeton University Press, 2021. 216 pp.

The book is not meant to be exhaustive, as Naish notes, and often focuses on the formative decades of the 1970s, '80s, and '90s, when discovery upon discovery raised questions about dinosaur social lives, parenting, physiology, and extinction, with each new proposal kicking off new debates that are running to this day. Little wonder it is known as the "Dinosaur Renaissance."

Illustrated with Naish's own drawings, the result is a solid primer on dinosaur science from *Archaeopteryx* to the Zigong Dinosaur Museum in China. Nevertheless, the book offers a view of modern dinosaur scientists that is practically petrified.

Naish includes profiles of a handful of paleontologists: Robert Bakker, Jack Horner, Halszka Osmólska, John Ostrom, Richard Owen, Greg Paul, and Paul Sereno. These figures were indeed pivotal in the dinosaur debates and discussions of the late 20th century, but Naish's decision to focus on them, rather than on contemporary paleontologists, makes the book feel decades out of date rather than representative of modern dinosaur studies. Aside from the gender imbalance, nonwhite scientists and researchers from the Global South are given short shrift.

Naish's book joins a number of recent titles that have failed to effectively convey the increasingly diverse practice of paleontology (*1*, *2*). I, too, have fallen far short in achieving equality and inclusivity in my writing (*3*).

There is no doubt that the Dinosaur Renaissance was huge for paleontology. Many of the children who were inspired by the museum exhibits, books, and films that debuted during that time are paleontologists or fossil fans still. But the height of that era's dinomania was nearly three decades ago, when discussions about diversity and representation in the field often occurred in the background, if they happened at all. Paleontologists today openly consider such issues, as well as adjacent topics such as the ethics of collecting specimens and samples in other countries and the repatriation of illegally exported fossils.

Still, there is much work to be done. In a field in which even gender equity between white cisgender researchers has been difficult to achieve, now is not the time to reaffirm the male-dominated days as representative of where the field stands today.

Change is likely to come slowly. Diversity in paleontology is currently highest among volunteers, students, and early-career researchers, all of whom are less likely to be thought of as authorities, less likely to be conducting research that is covered by the press, and less likely to write books themselves.

Again and again, op-eds and sociological studies have pointed out that a lack of visible representation affects who goes into science and who is supported through its process, which, in turn, affects scientific theory and thought (*4*–*7*). It is time to start embodying the change we wish to see. Ensuring that popular accounts of paleontology reflect the field's 21st-century practitioners would be a strong step toward this goal. ■

The reviewer is the author of *The Last Days of the Dinosaurs* (St. Martin's, forthcoming 2022) and *Skeleton Keys: The Secret Life of Bone* (Riverhead Books, 2019). Email: evogeek@gmail.com

REFERENCES AND NOTES

1. S. Brusatte, *The Rise and Fall of the Dinosaurs* (William Morrow, 2018).
2. A. Berta, S. Turner, *Rebels, Scholars, Explorers* (Johns Hopkins Univ. Press, 2020).
3. B. Switek, *My Beloved Brontosaurus* (Scientific American/Farrar, Straus and Giroux, 2013).
4. C. T. Begeny *et al.*, *Sci. Adv.* **6**, eaba7814 (2020).
5. K. Rainey *et al.*, *Int. J. STEM Educ.* **5**, 10 (2018).
6. K. Kricorian *et al.*, *Int. J. STEM Educ.* **7**, 16 (2020).
7. J. G. Stout *et al.*, *J. Pers. Soc. Psychol.* **100**, 255 (2011).

10.1126/science.abj2629

COVID-19

Dispatches from the front lines

A new film documents the fraught first days of the COVID-19 pandemic in New York City

By **Robert S. Krauss**



The First Wave follows health care workers and patients grappling with COVID-19 in the spring of 2020.

The opening shot of *The First Wave* is an aerial view of New York City's skyline enshrouded by a miasma of low-lying gray clouds. The ominous sensation conjured is apt, as the film returns viewers to March 2020, a time when doctors and nurses were faced with great unknowns as they cared for patients with COVID-19 and New York's hospitals were stretched to the breaking point. Between March and June, director Matthew Heineman and his team were granted extraordinary access to Northwell Health's Long Island Jewish Medical Center (LIJMC) in Queens, a location that was hit particularly hard at the outset of the pandemic. *The First Wave* brings viewers directly into the lives of LIJMC doctors, nurses, and patients as they navigated an acute public health crisis. The film is distressing, graphic, heartbreaking, and hopeful, and it is a must-see.

The First Wave proceeds chronologically, documenting, as they unfolded, the travails of health care teams, patients, and their families. Two patients hit very hard by COVID-19, Ahmed Ellis and Brussels Jabon, are followed at length, and their experiences provide a window into the jarring ups and downs encountered during the course of the disease and how care teams and families endured the associated uncertainty and anxiety.

The First Wave
Matthew Heineman, director
Neon, 2021.
93 minutes

Ellis, a Black 36-year-old school safety officer, was on a ventilator for a prolonged period of time. His recuperation was buoyed by steady and loving—albeit remote—communication with his wife and two young children and by the dedication of those who cared for him. Ellis develops a palpable bond with his physical therapist, Karl Arabian, who banters with him during his long recovery, eventually prompting Ellis to make jokes of his own.

Meanwhile, Jabon—a nurse originally from the Philippines—worked in the Northwell system and was 9 months pregnant when she contracted COVID-19. Her baby was delivered by emergency cesarean section just before Jabon was placed on a ventilator. Jabon lives in a multigenerational household, and her husband and parents are also nurses; to ensure the safety of her newborn, the family made the difficult decision to send him to live with cousins.

As Ellis and Jabon slowly recover, small things—the movement of a hand, a look of alertness—loom large, and the nursing staff develops hopeful customs that celebrate progress and ease tension, such as playing "Here Comes the Sun" whenever a patient comes off a ventilator.

Another central figure in the film is Nathalie Dougé, a first-generation Haitian American physician born and raised in the Bronx. Dougé's naturally upbeat presence is tested by the nearly inexorable deaths of patients whose families cannot be present at the hospital. She proves resilient in the face of the constraints of the time, taking joy in a surprise Zoom birthday party organized by family and friends. Nevertheless, the burden she shoulders is made heavier by the fact that people of color, many with backgrounds similar to her own, make up an outsized share of those hospitalized and dying.

As the emotional toll on Dougé mounts, the murder of George Floyd occurs, and New York's streets go from nearly deserted in March to full of protesters in May. She joins marches wearing a mask on which she has printed "I can't breathe," words spoken by both Mr. Floyd and many of her patients. Nearing her own emotional limits on two fronts, she still manages to bring her bedside manner to the streets, talking down a protester confronting a police officer before the altercation turns violent.

Despite their protestations to the contrary, the care teams at LIJMC were indeed heroic. When met with a scene of New Yorkers banging pots and pans at their windows in celebration of health care workers, viewers of this documentary may feel like joining in anew. Such moments provide relief from the film's unflinching and intimate look at death, suffering, and grief.

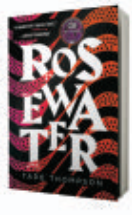
The film closes with the same imagery with which it opened—the Manhattan skyline cloaked in a miasma. Notwithstanding the rapid development of highly effective vaccines in the interim, this closing vista reminds viewers that the COVID-19 pandemic is still very much with us. Watching *The First Wave* also reminds us what is at stake. ■

10.1126/science.abn2200

The reviewer is at the Department of Cell, Developmental, and Regenerative Biology, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA. Email: robert.krauss@mssm.edu

PHOTO: NATIONAL GEOGRAPHIC

SCIENCE & RACE



PODCAST
Rosewater
Tade Thompson
Orbit, 2018.
432 pp.

The idea of race is rooted in historical conceptions about identity and human difference. It has changed over time and is likely to continue to change. In the final episode of our series on science and race, Tade Thompson discusses his novel *Rosewater*, which imagines a future society with a radically new understanding of humanity that is still grappling with the cultural legacies of racism and colonialism.
https://scim.ag/3EGErwo

10.1126/science.abn0868



Many vultures, including Andean condors like this one, have recently been found poisoned.

PHOTO: FACUNDO VITAL

LETTERS

Edited by **Jennifer Sills**

Governments must halt vulture poisoning

Poisoning is currently the most pervasive threat to vultures (*1*). Despite efforts to better protect these birds [e.g., (*2*, *3*)], the problem seems to have increased in intensity over the past 2 years. Between September 2019 and March 2020, more than 2000 Critically Endangered hooded vultures (*Necrosyrtes monachus*) were poisoned with methiocarb across eastern Guinea-Bissau (*4*), representing the death of approximately 4.6% of the population of this species in this country (*5*). Similar to the worst known case of Andean condor (*Vultur gryphus*) poisoning (34 individuals in Argentina in 2018) (*1*), 34 condors were killed by pesticides in Bolivia this year (*6*). Poisoning incidents were also worrying in many other regions of the world, such as Nepal, where 67 vultures were poisoned in April 2021 (*7*); Spain, where nearly 100 vultures were poisoned in June 2021 (*8*); and Israel, where 5% (9 of 200) of the population of Griffon vultures (*Gyps fulvus*) from this country were killed in just one incident with pesticides in October (*9*). Governments must take immediate action to mitigate harm to vultures (*10*).

Reasons for vulture poisoning are diverse. Some vultures are unintentional victims of conflict between mammalian predators and humans (*1*). Others are poisoned intentionally by humans who believe that vultures are the main culprits behind livestock losses (*1, 11*). In some cases, vultures are killed to provide parts for traditional medicine (*4*). Those who gather near an illegally killed animal are sometimes poisoned by poachers in an effort to escape detection (*10*). To reduce poisoning, we must increase tolerance and empathy toward these birds and other wildlife. Highlighting the contributions vultures make to nature and people and the potential consequences of their extinction could help achieve these goals (*11*).

Preventing poisoning events and analyzing poisoned carcasses can be expensive and very difficult in some countries. Governments of every country inhabited by vultures must commit to their protection and allocate more funds to solve this conservation problem. They will have to work collaboratively considering vulture home ranges may include several countries. Authorities should also promote anti-poisoning measures: education, legislation (such as traceability of pesticides), rapid diagnosis of poisoning events, support for local communities, and in-depth study of human–vulture conflict. We must halt wildlife poisoning given that we may be reaching our last opportunity to prevent the extinction of some threatened vultures.

Pablo I. Plaza* and Sergio. A. Lambertucci
Grupo de Investigaciones en Biología de la Conservación, Instituto en Biodiversidad y Medio Ambiente, Universidad Nacional del Comahue - CONICET, Quintral 1250 (R8400FRF), Bariloche, Argentina.
*Corresponding author.
Email: plazapablo@comahue-conicet.gob.ar

REFERENCES AND NOTES

1. P. I. Plaza, E. Martínez-López, S. A. Lambertucci, *Sci. Tot. Environ.* **687**, 1207 (2019).
2. A. J. Botha *et al.*, "Multi-species action plan to conserve African-Eurasian vultures," CMS Raptors MOU Technical Publication No. 5 (Coordinating Unit of the CMS Raptors MOU, Abu Dhabi, United Arab Emirates, 2017).
3. P. I. Plaza, S. A. Lambertucci, *Biol. Conserv.* **256**, 108924 (2021).
4. M. Henriques *et al.*, *Science* **370**, 304 (2020).
5. M. Henriques *et al.*, *PLOS ONE* **13**, e0190594 (2018).
6. Y. S. Praeli, "Amenaza latente: 34 cóndores mueren envenenados en Bolivia" (2021); https://es.mongabay.com/2021/02/condor-andino-envenenamiento-bolivia/ [in Spanish].
7. "Significant blow to vulture conservation: Largest mass vulture mortality event reported from Nepal severely impacts Endangered Species," Vulture Conservation Foundation (2021); https://4vultures.org/blog/significant-blow-to-vulture-conservation-largest-mass-vulture-mortality-event-reported-from-nepal-severely-impacts-endangered-species/.
8. "Dead or dying: Nearly 100 vultures fall victim to suspected mass poisoning in Castilla y León (Spain)," Vulture Conservation Foundation (2021); https://4vultures.org/blog/dead-or-dying-nearly-100-vultures-fall-victim-to-suspected-mass-poisoning-in-castilla-y-leon-spain/.
9. "Poisoned! Mass mortality event of Griffon Vultures in Israel causes significant blow to vulnerable population," Vulture Conservation Foundation (2021); https://4vultures.org/blog/poisoned-mass-mortality-event-of-griffon-vultures-in-israel-causes-a-significant-blow-to-the-vulnerable-population/.
10. A. Margalida, D. Ogada, A. Botha, *Science* **365**, 1089 (2019).
11. S. A. Lambertucci *et al.*, *Conserv. Sci. Pract.* **3**, e415 (2021).

10.1126/science.abn2066

Transgender rights rely on inclusive language

Inclusive language around sex diversity has never been more important. In the 5 years since the National Institutes of Health's notice that proposals address sex as a biological variable (*1*), the sociopolitical climate has become dangerously hostile to transgender and gender-diverse people. In the United States, a record number of anti-trans bills were proposed this year, and gender-affirming healthcare faces record wait times and unprecedented legal challenges in the United Kingdom (*2*). Science can seem far removed from the struggle for transgender equality, but scientific language directly affects trans-focused rhetoric and policy (*3*–*9*).

It is important to recognize the context-dependent and multidimensional nature of sex. Rather than privilege any characteristic as the sole determinant of sex, "male" and "female" should be treated

as context-dependent categories with flexible associations to multiple variables (such as, but not limited to, genitalia, gametes, or karyotype). The usage of "male" and "female" should be explicitly defined in any given study (*10*). Failing to do so promotes harmful language (such as "male chromosomes" rather than "Y chromosomes") that attributes an essential "maleness" or "femaleness" to traits, obscuring the true biological mechanisms at work (e.g., the *Tdf* gene leads to testicular development, not to "being male"). No one trait determines whether a person is male or female, and no person's sex can be meaningfully prescribed by any single variable.

Awareness of the distinction between sex and gender is another vital element to inclusive, quality research (*11*). Conflating the two harms and invalidates gender minorities by implying that these distinct attributes are inextricably linked. It is in this context that many US states require surgery or sterilization before a person can change their gender on legal documents. We have a responsibility to use precise language both as researchers studying sex-associated variables and as members of a diverse academic community.

As scientists, we must push back against the misappropriation of biological terms by promoting precise language that focuses on the variables themselves (e.g., "menstruating people") and acknowledging that people express these variables in ways that may not conform with a binary system of sex or gender. This both creates a more inclusive environment for gender-diverse scientists and reinforces that sex is a context-dependent summary of a multidimensional variable space.

Miriam Miyagi[1]*, Eartha Mae Guthman[2,3], Simón(e) Dow-Kuang Sun[3,4]
[1]Harvard University, Cambridge, MA 02138, USA. [2]Princeton University, Princeton, NJ 08544, USA. [3]Center for Applied Transgender Studies, Chicago, IL 60657, USA. [4]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA.
*Corresponding author.
Email: miriam.miyagi@gmail.com

REFERENCES AND NOTES

1. M. E. Arnegard *et al*., *J. Wom. Health* **29**, 858 (2020).
2. S. Rosenthal, *Nat. Rev. Endocrinol*. **17**, 581 (2021).
3. F. Ashley, *Sociological Rev*. **68**, 779 (2020).
4. N. Q. Ha *et al*., *Arch. Sex. Behav*. **43**, 1035 (2014).
5. S. Hines, *Sociological Rev*. **68**, 699 (2020).
6. E. A. Sharrow, *Sports Transgender Rights Bodily Polit. Cisgender Supremacy Laws* **10**, 63 (2021).
7. M. Sudai, *Harvard J. Law Gender* **421**, 460 (2019).
8. J. Wuest, *Polit. Gender* **15**, 336 (2019).
9. J. Wuest, *Perspect. Polit*. **19**, 838 (2020).
10. S. Richardson, *Philos. Theory Pract. Biol*. **13**, 10.3998/ptpbio.16039257.0013.009 (2021).
11. Y. Garcia-Sifuentes, D. L. Maney, *eLife* **10**, e70817 (2021).

10.1126/science.abn3759

Brazil's deception threatens climate goals

Brazilian agribusiness has led a lobby responsible for dismantling the country's environmental and Indigenous legislation (*1*). In the lead-up to the United Nations Climate Change Conference (COP26), Brazil began proposing policies that were misleadingly presented as beneficial to the environment. The proposals made by Brazil's Minister of Environment at COP26 followed the same pattern (*2*).

In the prelude to COP26, a bill (PL 528/2021) supported by the "ruralist" (large landholder) lobby advanced through the National Congress to create the "Brazilian Market for Emissions Reductions." The proposed Brazilian market is not primarily to invest in reducing deforestation but rather for measures such as paying soy planters to use no-till methods, a subsidy that has no climate benefits. [These methods are less expensive and therefore would be adopted even without subsidies (*3*)]. The market would also subsidize planting trees in deforested areas, in an effort to recover the "areas of permanent protection" ("*áreas de proteção permanente*" in the Brazilian Forest Code) that ranchers have illegally cleared (*4*). Luckily, the bill was put on hold on 9 November (*5*); if enacted, as is likely, it would have limited benefits for mitigating global warming. Effective policies would instead prevent deforestation, which would protect the water cycle, biodiversity, and the forest's traditional peoples. Brazil's environment could have much to gain from a carbon market (*6*), but the country would be better off without this bill (*7*).

Three weeks before COP26, Brazil's presidential administration launched its "green growth program." The program contains no specific measures, numerical targets, or timelines (*8*). Brazil's main promise is to achieve "zero illegal deforestation," promised by 2030 in the Paris Agreement and again at COP26. However, this can be achieved either by halting deforestation or by simply legalizing the clearing that is taking place, and Brazil is opting for the latter (*9*).

In the country's COP26 presentation, Brazil altered its greenhouse-gas emission accounts by switching to a higher estimate for emissions in the base year before calculating its promised percentage emission reduction (*10*). In addition, Brazil's chosen base year (2005) misleads because it is near the notorious 2004 deforestation peak (*10*). This makes Brazil's promise to reduce its emissions by 50% by 2030, instead of just 43%, meaningless (*10*). Furthermore, the Brazilian COP26 proposal demands payment from other countries for Amazonia's environmental services (*4*). Yet the payment would apparently simply supplement the national budget rather than be earmarked for environmental actions (*4*).

Brazil's current stance on environmental issues should be seen as a global threat. Various legal paths exist to oblige compliance with the "non-binding" promises made at COP26 (*11*). Brazil can also be induced to comply by the countries that import Brazilian commodities, especially China (*12*). Given the threat that Brazilian agribusiness poses to global efforts to contain climate change, importing countries should condition their purchases on annual reductions in deforestation and emissions being in line with the COP26 promises.

Lucas Ferrante* and Philip M. Fearnside
National Institute for Research in the Amazon (INPA), Ecology Graduate Program, Manaus, AM, Brazil.
*Corresponding author.
Email: lucasferrante@hotmail.com

REFERENCES AND NOTES

1. L. M. M. Vale *et al*., *Biol. Conserv.* **255**, 108994 (2021).
2. C. Angelo, F. Werneck, N. Afonso, S. Watanabe Jr., "Na COP26, ministro mente sobre ambição de meta climática do Brasil," *O Eco* (2021); https://oeco.org.br/reportagens/na-cop26-ministro-mente-sobre-ambicao-de-meta-climatica-do-brasil/ [in Portuguese].
3. A. Ofstehage, R. Nehring, *Int. Journ. Agric.l Sustain.* **19**, 335 (2021).
4. P. M. Fearnside, "COP-26: Amazonia's Indigenous peoples are vital to fighting global warming (commentary)," *Mongabay* (2021); https://news.mongabay.com/2021/11/cop-26-amazonias-indigenous-peoples-are-vital-to-fighting-global-warming-commentary/.
5. A. C. Amaral, "Governo Bolsonaro barra votação sobre mercado de carbono na Câmara," *Folha de São Paulo* (2021); https://www1.folha.uol.com.br/ambiente/2021/11/governo-bolsonaro-barra-votacao-sobre-mercado-de-carbono-na-camara.shtml [in Portuguese].
6. P. M. Fearnside, *Clim. Change* **118**, 509 (2013).
7. Grupo Carta de Belém, "Com tramitação a toque de caixa, PL 528 quer regular Mercado Brasileiro de Redução de Emissões" (2021); www.cartadebelem.org.br/com-tramitacao-a-toque-de-caixa-pl-528-quer-regular-mercado-brasileiro-de-reducao-de-emissoes/ [in Portuguese].
8. B. Caram, M. Vargas, "Governo prepara 'programa de crescimento verde' para investidores, mas Economia pediu regras ambientais frouxas," *Folha de São Paulo* (2021); https://www1.folha.uol.com.br/mercado/2021/09/governo-prepara-programa-de-crescimento-verde-para-apresentar-a-investidores.shtml [in Portuguese].
9. L. Ferrante *et al*., *Land Use Pol.* **108**, 105559 (2021).
10. Observatório do Clima, "Nova meta do Brasil no clima reduz pedalada, mas não a Vergonha" (2021); www.oc.eco.br/en/nova-meta-do-brasil-no-clima-reduz-pedalada-mas-nao-a-vergonha/ [in Portuguese].
11. J. Catanos, "COP26: Are climate declarations and emission reduction pledges legally binding?" *Mongabay* (2021); https://news.mongabay.com/2021/11/cop26-are-climate-declarations-and-emission-reduction-pledges-legally-binding/.
12. L. Kehoe *et al*., *Science* **364**, 341 (2019).

10.1126/science.abn0693



A vaccination center in Mexico, a country that was the focus of one award winner.

PHOTO: PEDRO PARDO/AFP/GETTY IMAGES

2021 AAAS Kavli Science Journalism Awards announced

Reports from the ongoing pandemic highlighted the entries

By **Earl Lane** *and* **Emily Hughes**

Six entries dealing with aspects of the continuing global COVID-19 pandemic are among the winners of the 2021 AAAS Kavli Science Journalism Awards. Other winners dealt with important issues of equity and ethics in the conduct of science.

The international awards, administered by the American Association for the Advancement of Science, recognize distinguished science reporting for a general audience. A Gold Award ($5000) and a Silver Award ($3500) are presented in each of eight categories.

The pandemic-related winners include a widely shared multimedia piece by Mariano Zafra and Javier Salas for Madrid's *El País* on how readily the coronavirus that causes COVID-19 can spread via aerosols in a living room, a bar, and a classroom; a piece for *Wired* by Megan Molteni on the historically flawed science behind the World Health Organization's initial guidance that the coronavirus spreads primarily via droplets from coughing rather than exhaled aerosol particles; and a sensitive account by Maartje Bakker of Amsterdam's *De Volkskrant* on the life—and death—of two rhesus monkeys used in research to develop a COVID-19 vaccine.

In a winning video entry, NOVA producer Arlo Pérez Esquivel told of leaving his home in Boston to stay with family in Uruapan, Mexico, because he felt he would be safer from COVID-19. Instead, he soon realized that people were dying all around him and that cases of the disease were being seriously underreported in Mexico.

Among entries that delved into the equity and conduct of science, "Picture a Scientist," a documentary produced for NOVA on PBS, won for its exploration of long-standing patterns of discrimination against women in science and the determined efforts of biologist Nancy Hopkins, chemist Raychelle Burks, and geologist Jane Willenbring to help change that.

CBC/Radio-Canada's science program "Quirks & Quarks" won for a special on the past and future of Black people in the sciences, including the history of biased and false "race science" that led to their mistreatment and misunderstanding by the scientific and medical communities.

Richard Van Noorden of *Nature* was honored for a piece on ethical questions surrounding the use of facial-recognition research, including the use of trained algorithms to distinguish faces of predominantly Muslim Uyghur people in China.

Winning entries also explored the beauty and mysteries of nature. Aathira Perinchery, writing for "FiftyTwo"—a digital publication based in India—explored deep questions about evolutionary biology while reporting on the diversity of new species being discovered in India. She is the first Indian journalist to win the AAAS Kavli award since the contest went global in 2015. Zafra and Salas also are the first journalists at a Spanish outlet to win the award.

Tony Bartelme of *The Post and Courier* of Charleston, S.C., won his second AAAS Kavli award for a story about an elusive marsh-dwelling bird called the eastern black rail and one researcher's efforts to study and protect it. Michael Werner also won his second AAAS Kavli award as part of a video team that produced segments for PBS on endangered prairie and forest habitats. Stephen Ornes won for the second time in the Children's Science News category.

"Congratulations to the talented journalists who enlightened their audiences about the science of the global pandemic, about efforts to diversify and strengthen the research enterprise, and about the beauty and complexity of the natural world," said Sudip Parikh, CEO of AAAS and executive publisher of the *Science* family of journals. "These awards continue to spotlight the value of informed, thought-provoking journalism."

Independent panels of science journalists select the winners. The awards program, endowed by The Kavli Foundation, drew entries from 47 countries in this year's contest. The winners will receive their awards in a virtual ceremony held in conjunction with the 2022 AAAS Annual Meeting in February.

2021 AAAS Kavli Science Journalism Award winners

Read a full description of the winning entries at https://sjawards.aaas.org.

Science Reporting, Large Outlet

Gold Award: Mariano Zafra and Javier Salas, *El País* (Spain), "A room, a bar and a classroom: How the coronavirus is spread through the air," Oct. 29, 2020

Silver Award: Roxanne Khamsi, Elemental (from Medium), "The Mystery Of Why Some People Keep Testing Positive For Covid-19," July 28, 2020

Science Reporting, Small Outlet

Gold Award: Aathira Perinchery, FiftyTwo (India), "Succession," Jan. 29, 2021

Silver Award: Tony Bartelme, *The Post and Courier* (Charleston, S.C.), "Ghost Bird," Sept. 13, 2020

Science Reporting – In-Depth (more than 5,000 words)

Gold Award: Noah Gallagher Shannon, *The New York Times Magazine*, "What's Going on Inside the Fearsome Thunderstorms of Córdoba Province?," July 22, 2020

Silver Award: Maartje Bakker, *De Volkskrant* (The Netherlands), "Hoe Chips en Dip stierven voor een vaccin" (How Chips and Dip died for a vaccine), July 10, 2021

Magazine

Gold Award: Megan Molteni, *Wired*, "The 60-Year-Old Scientific Screwup That Helped Covid Kill," May 13, 2021

Silver Award: Richard Van Noorden, *Nature*, "The ethical questions that haunt facial-recognition research," Nov. 18, 2020

Video

Spot News/Feature Reporting (20 minutes or less)
Gold Award: Michael Werner, Joe Hanson, Rachel Raney, and Brandon Arolfo, PBS, "How Bison Are Saving America's Lost Prairie", Jan. 14, 2021, "Inside the Fight to Save an Ancient Forest (and the Secrets it Holds)," July 1, 2021

Silver Award: Arlo Pérez Esquivel, NOVA/GBH for PBS, "Mexico's COVID Cases and Deaths are Underreported—Why?," April 2, 2021

In-Depth Reporting (more than 20 minutes)

Gold Award: Sharon Shattuck, Ian Cheney, Manette Pottle, and Amy Brand, A NOVA Production by Uprising LLC for GBH Boston, "Picture a Scientist," April 14, 2021

Silver Award: Adam Bolt, Regina Sobel, Elliot Kirschner, Sarah Goodwin, and Meredith DeSalazar, A NOVA Production by the Wonder Collaborative for GBH Boston, "Human Nature," Sept. 9, 2020

Audio

Gold Award: Amanda Buckiewicz and Nicole Mortillaro, CBC/Radio-Canada, "Quirks & Quarks: Black in Science special," Feb. 27, 2021

Silver Award: Fiona Broom, Charles Pensulo, Jubiel Zulu, Brook Abdu, and Harrison Lewis, SciDev.Net (United Kingdom), "Rising COVID-19 cases and vaccine doubt", Feb. 3, 2021, "Debunking COVID-19 myths and remedies," Feb. 10, 2021 "Science and you: Africa's COVID-19 vaccines," Feb. 17, 2021

Children's Science News

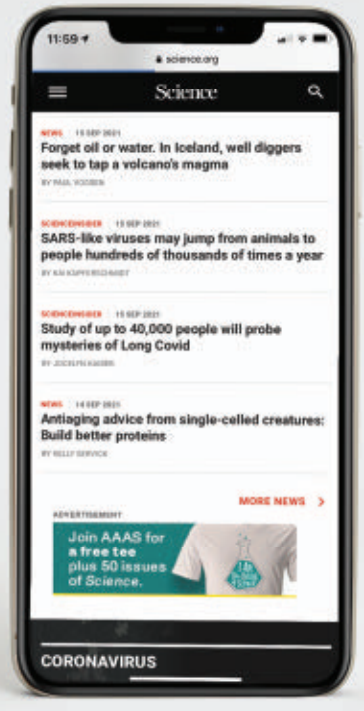
Gold Award: Bridgett Henwood, Estelle Caswell, Kimberly Mas, and Elizabeth Scheltens, Vox, "Why bird nests aren't covered in poop," Sept. 2, 2020, "The secret history of dirt", Sept. 3, 2020, "How to be a Cloud Detective," Sept. 4, 2020

Silver Award: Stephen Ornes, Science News for Students, "Whales get a second life as deep-sea buffets," Oct. 15, 2020

NEWS FROM
Science

Up-to-the-minute research and policy news you won't find in print

Visit us online to read all the news coverage that there just wasn't enough room to print in this issue.



Science.org/news

RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**



A cryo–electron microscopy map reveals a molecule of RNA (red) being transferred from one polymerase (green) to the next (purple) through a protein channel.

MOLECULAR BIOLOGY

A tunnel between RNA polymerases

Eukaryotes encode five multiple-subunit, DNA-dependent RNA polymerases, of which Pol I, Pol II, and Pol III function as single units to produce cellular single-stranded RNA. The plant-specific Pol IV forms a complex with RDR2 (an RNA-dependent RNA polymerase) to produce double-stranded precursors of small interfering RNA essential for genomic DNA methylation. Huang *et al*. determined the cryo–electron microscopy structures of the Pol IV-RDR2 complex. The structures show that Pol IV and RDR2 connect their active centers through an inner RNA transfer channel and that Pol IV reverses transcription direction and hands over its transcript directly through the channel to RDR2 for the production of the second strand of the double-stranded RNA. —DJ *Science*, abj9184, this issue p. 1579

ZEOLITES

Extra-large and intersecting pores

Thermally stable zeolites with large pores, such as zeolite Y with 12-membered rings, are used for converting large molecules from petroleum into smaller, more useful hydrocarbons. Lin *et al*. report the synthesis of ZEO-1, a thermally stable zeolite with extra-large, 16-membered rings. The use of tricyclohexylmethylphosphonium as an organic structure–directing agent created a zeolite with a fully interconnected multidimensional framework and a very high specific surface area. —PDS

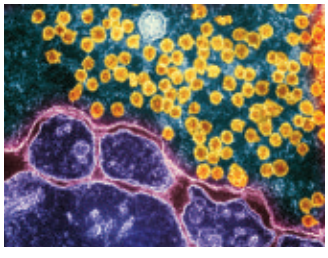
Science, abk3258, this issue p. 1605

NEURONAL SIGNALING

Gβγ dimers get specific at the synapse

In contrast to how much is known about G protein–α subunits, comparatively little is known about G protein–βγ dimers, which regulate synaptic transmission by inhibiting the release of neurotransmitter-containing vesicles from presynaptic neurons. Yim *et al*. identified the Gβ and Gγ subunits that interacted with the SNARE complex, which mediates vesicle exocytosis. This analysis revealed that the type of Gβ and Gγ subunits that interacted with SNARE proteins changed with epinephrine stimulation and depended on the type of neuron. —JFF

Sci. Signal. **14**, eabc4970 (2021).



False-color transmission electron microscopy image of an excitatory synapse showing vesicles primed for exocytosis

CORONAVIRUS

Path to another drug against COVID-19

The rapid development of vaccines has been crucial in battling the ongoing COVID-19 pandemic. However, access challenges remain, breakthrough infections occur, and emerging variants present increased risk. Developing antiviral therapeutics is therefore a high priority for the treatment of COVID-19. Some drug candidates in clinical trials act against the viral RNA-dependent RNA polymerase, but there are other viral enzymes that have been considered good targets for inhibition by drugs. Owen *et al*. report the discovery and characterization of a drug against the main protease involved in the cleavage of polyproteins involved in viral replication. The drug, PF-07321332, can be administered orally, has good selectivity and safety profiles, and protected against infection in a mouse model. In a phase 1 clinical trial,

CREDITS: (TOP TO BOTTOM) HUANG *ET AL*.; DENNIS KUNKEL/MICROSCOPY/SCIENCE SOURCE

the drug reached concentrations expected to inhibit the virus based on in vitro studies. It also inhibited other coronaviruses, including severe acute respiratory syndrome coronavirus 1 and Middle East respiratory syndrome coronavirus, and could be in the armory against future viral threats. —VV

Science, abl4784, this issue p. 1586

ELECTROCHEMISTRY

Boosting ammonia with a little oxygen

Ammonia synthesis from nitrogen for fertilizer production is highly energy intensive. Chemists are therefore exploring electrochemical approaches that could draw power from renewable sources while generating less waste. One promising cycle involves the reduction of lithium ions at an electrode, with the resultant metal in turn reducing nitrogen and regenerating the ions. Li *et al*. report the counterintuitive result that small quantities of oxygen could enhance the efficiency of this process, which they attribute to diffusional effects that limit excessive lithium reduction. —JSY

Science, abl4300, this issue p. 1593

IMMUNOTHERAPY

Another benefit of dietary fiber

The gut microbiome can modulate the immune system and influence the therapeutic response of cancer patients, yet the mechanisms underlying the effects of microbiota are presently unclear. Spencer *et al*. add to our understanding of how dietary habits affect microbiota and clinical outcomes to immunotherapy. In an observational study, the researchers found that melanoma patients reporting high fiber (prebiotic) consumption had a better response to checkpoint inhibitor immunotherapy compared with those patients reporting a low-fiber diet. The most marked benefit was observed for those patients reporting a combination of high fiber consumption and no use of over-the-counter probiotic supplements. These findings provide early insights as to how diet-related factors may influence the immune response. —PNK

Science, aaz7015, this issue p. 1632

CARBON NANOTUBES

Straining to make a transistor

The use of carbon nanotubes (CNTs) as short-channel-length transistors will require control of their chirality, which determines whether they are semiconducting or metallic and if they form strong, low-resistance contacts. Tang *et al*. fabricated CNT intramolecular transistors by progressive heating and straining of individual CNTs within a transmission electron microscope. Changes to chirality along sections of the nanotube created metallic-to-semiconducting transitions. A semiconducting nanotube channel was covalently bonded to the metallic nanotube source and drain regions. The resulting CNT intramolecular transistors had channel lengths as short as 2.8 nanometers. —PDS

Science, abi8884, this issue p. 1616

ANTIBIOTICS

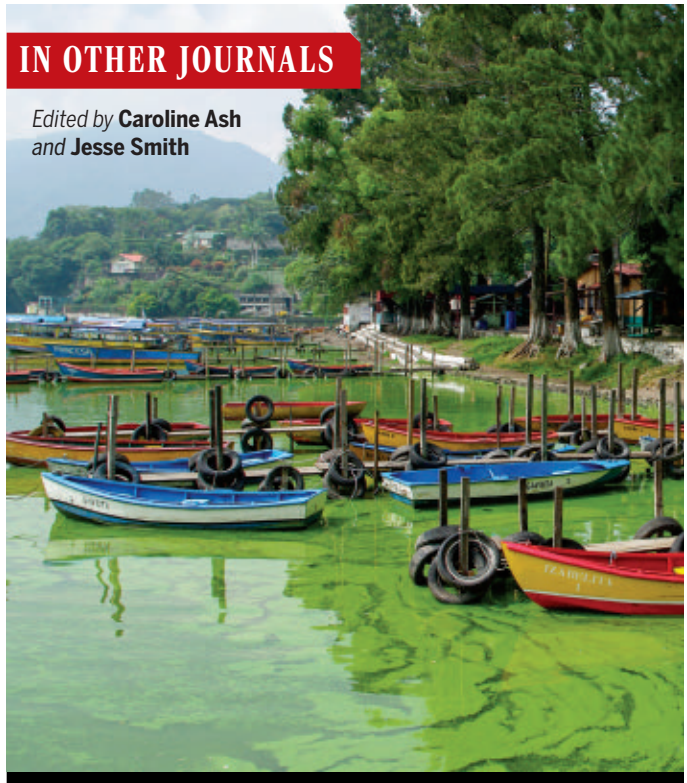
Metabolite mitigates resistance

Antibiotic resistance is a pervasive and increasing problem. Zhao *et al*. show that glutamine promotes increase the influx and efficacy of a variety of antibiotics against multiple species of pathogenic Gram-negative bacteria. Mechanistically, glutamine promoted changes in nucleotide metabolism that led to the upregulation of a two-component regulatory system, resulting in increased, nonspecific membrane permeability in bacteria having varied resistance mechanisms. Glutamine supplementation was also effective in mouse models of systemic or biofilm infection, suggesting a potential strategy to enhance antibiotic efficacy. —CAC

Sci. Transl. Med. **13**, eabj0716 (2021).

IN OTHER JOURNALS

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



PALEOLIMNOLOGY

Pollution in prehistory

Harmful algal blooms in oceans and lakes are a frequent consequence of anthropogenic pollution. For instance, sewage discharge or nutrient runoff from farms can lead to anoxic conditions toxic to eukaryotic life. Waters *et al.* show that this is not just a phenomenon of the industrial age. Sediment cores from Lake Amatitlán, Guatemala, indicate the presence of cyanobacteria characteristic of harmful blooms during the pre-Columbian Mayan occupation of the watershed, with algal spikes coinciding with periods of maximum human population. At their peak, chemical signatures of the algal blooms in Mayan times show concentrations and toxicity similar to present-day blooms in the same lake. —AMS

Proc. Natl. Acad. Sci. U.S.A. **118**, e2109919118 (2021).

Algal blooms occur in modern-day Lake Amatitlán as result of human activities, as they did in prehistoric times.

HIV

T_{RM} cells are state of the ART

Recent advances in antiretroviral therapy (ART) and testing have restored the life expectancies of people infected with HIV. However, HIV-positive individuals still show a higher risk for malignancies of the skin and mucosa associated with human papilloma virus (HPV). Saluzzo *et al.* report that HIV-positive individuals who begin ART late show good recovery of circulating CD4 T cells but exhibit an irreversible depletion of CXCR3$^+$ tissue-resident memory T (T_{RM}) cells in the skin. Their skin is instead repopulated by T helper 2–like cells, which appear to foster a permissive immune environment. By contrast,

CREDIT: MAX HERMAN/ALAMY STOCK PHOTO



GENETICS

Single-sex offspring

Agriculture and research practices often favor one sex over the other. For egg production, female chickens are needed, whereas for beef production, males are preferred. To eliminate the need to cull animals, methods are being explored to reliably generate single-sex populations. Douglas *et al.* succeeded in this goal by using a CRISPR-Cas9 strategy to efficiently produce only male or female mice. When males carrying a Cas9-encoding transgene on either the X or Y chromosome are mated with females with a guide RNA that targets the topoisomerase 1 gene, only one sex survives to birth. It is likely that this method can be applied to other vertebrate species. Although this technique solves the ethical issues associated with the culling of unwanted livestock, it could introduce other ethical concerns. —BAP *Nat. Commun.* **12**, 6926 (2021).

New techniques may help agriculture facilities, such as this large-scale egg farm, obtain animals of a single sex without culling.

HIV-positive individuals who go on ART shortly after infection reconstitute their cutaneous T_{RM} cells within a year. Thus, early diagnosis and treatment of HIV infection may help to stave off HPV-related cancers. —STS

Immunity 10.1016/j.immuni.2021.10.021(2021).

CANCER

A new cancer target?

Therapeutic treatments for metastatic cancer are an important unmet need. Studies have shown that the interaction between metadherin (MTDH) and staphylococcal nuclease domain-containing (1SND1) is important in tumor initiation, progression, and metastasis in mice. Shen *et al*. show that deletion of the *Mtdh* gene in established mammary tumors in mice suppressed tumor development, reduced metastasis, and improved survival. This effect was dependent on their interaction with SND1. The authors also found compounds that could disrupt the interaction by binding to SND1. Their lead molecule reduces tumor growth and lung metastasis in established mammary tumors in vivo. Preventing the MTDH–SND1 interaction also sensitizes mammary tumors to chemotherapy. This indicates that interfering with this protein-protein interaction could be a viable strategy for clinical investigation. —GKA

Nat. Cancer 10.1038/s43018-021-00279-5 (2021).

POROUS FRAMEWORKS

Chalcogen-bonded organic frameworks

The assembly of porous supramolecular organic frameworks is often achieved by designing hydrogen-bonding interactions, but the use of other noncovalent interactions could expand the types of materials that could be synthesized. Eckstein *et al*. derivatized the ends of the propeller-shaped molecule triptycene with aromatic rings containing selenium and nitrogen atoms to form triptycene tris(1,2,5-selenadiazole). In nonpolar solvents, this molecule assembles through nitrogen lone-pair electrons donating into selenium s holes. The resulting antiparallel motifs form ~2-nanometer-diameter hexagonal channels. These frameworks can undergo host-guest exchange, such as replacing 1,2,5-trichlorobenzene with 1-chloronaphthalene. The topology of the frameworks is sensitive to solvent, and the one grown from 1-chloronaphthalene was altered after exposure to carbon tetrachloride vapor. —PDS

J. Am. Chem. Soc. **143**, 20207 (2021).

MACHINE LEARNING

Learning about new classes of antibiotics

Machine learning (ML) is increasingly being used as an advisory tool to guide antibacterial drug searches. Such methods typically use large training datasets and require extensive prior study of structure-activity relationships. ML models have proven to be effective within previously studied classes of antibiotics, but their predictive power is significantly limited when the task is to find candidate molecules that do not resemble known antibiotics. Tiihonen *et al.* present a general ML framework that does not rely on prior knowledge of the mechanism of antibiotic action. Their ML model developed for conjugated oligoelectrolyte molecules successfully predicted their antimicrobial activity against *E. coli* K12 and could potentially be applied to other understudied antibiotic classes. —YS

J. Am. Chem. Soc. 10.1021/jacs.1c05055 (2021).

PHOTOVOLTAICS

A solar cell doubling up as a battery

Solar power is hampered by daily and seasonal fluctuations of power output. This inconsistency can be compensated for by using energy storage such as rechargeable batteries but will increase the total cost and decrease the overall efficiency of the system. Jiang *et al.* developed a solar-powered battery in which the coexistence of photoelectric conversion and storage in a single device is achieved for the first time. The device consists of a van der Waals heterojunction between a tungsten diselenide layer and a quasi–two-dimensional electron gas, has an external quantum efficiency of more than 90%, and can retain its charge for up to 7 days, but requires the liquid nitrogen temperature to be functional. —YY

Phys. Rev. Lett. **127**, 217401 (2021).

PHOTO: ALEXIS LLORET/ALAMY STOCK PHOTO

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

CORONAVIRUS

Vaccine breakthrough infections

The vaccines developed against severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) confer very high protection against severe COVID-19. However, as the vaccine rollout has progressed and variants have emerged, it has become clear that immune protection can wane 4 to 6 months after the second immunization, especially in older people and those who are immunocompromised. With vaccine breakthrough infections increasing, particularly with the prevalence of the Delta variant, impressive responses have been obtained from booster immunizations. In a Perspective, Gupta and Topol discuss the emergence of breakthrough infections and the need for booster doses to improve immune protection. Although boosters are important, so are nonpharmacological interventions such as masking, which help to reduce cases of COVID-19 and thereby limit the opportunity for the emergence of variants that can evade immunity. —GKA

Science, abl8487, this issue p. 1561

IMMUNOLOGY

Fibrin gums up the works

Plasmin is an abundant plasma protease that cleaves and deactivates the clot-associated protein fibrin. Human deficiencies in plasmin and its inactive proenzyme form, plasminogen (PLG), cause severe inflammation in mucosal tissues such as the mouth and eyes. Silva *et al*. report that, like humans, mice lacking plasminogen accumulate extravascular fibrin and develop an oral pathology that phenocopies human ligneous periodontitis (see the Perspective by Vicanolo and Hidalgo). The excess fibrin activates neutrophils through the αMβ2 (Mac-1) integrin receptor, which triggers the production of reactive oxygen species and neutrophil extracellular traps. Additionally, certain human polymorphisms in the PLG gene were found to be associated with increased likelihood of developing periodontitis, suggesting that fibrin–neutrophil interactions may be an attractive target for future treatments of this prevalent disease. —STS

Science, abl5450, this issue p. 1575; see also abn0399, p. 1559

ENERGETICS

Efficiency leads to leisure

Humans are animals—merely another lineage of great apes. However, we have diverged in significant ways from our ape cousins and we are perennially interested in how this happened. Kraft *et al*. looked at energy intake and expenditure in modern hunter-gatherer societies and great apes. They found that we do not spend less energy while foraging or farming, but we do acquire more energy and at a faster rate than our ape cousins. This difference may have allowed our ancestors to spend more time in contexts that facilitated social learning and cultural development. —SNV

Science, abf0130, this issue p. 1576

PLANT SCIENCE

Root meristem controls

The plant meristem, a small cluster of stem cells, generates all of the cell types necessary for the plant's indeterminate growth pattern. Roszak *et al*. use single-cell analyses to follow development from the stem cell to the enucleated cell of the phloem vasculature. In the root of the small mustard plant *Arabidopsis*, this process takes just over 3 days, and the developmental trajectory spans more than a dozen different cell states. A transcriptional program initially held under repressive control is released as those initial repressors dissipate. Reciprocal repression by regulators early and late in the developmental trajectory control a rapid switch in the differentiation program. —PJH

Science, aba5531, this issue p. 1577

PALEONTOLOGY

Early marine giant

The largest animals to have ever lived occupied the marine environment. Modern cetaceans evolved their large size over tens of millions of years in response to the increased productivity of cold marine waters. However, whales were not the first marine giants to evolve. Sander *et al*. describe a 244-million-year-old fossil ichthyosaur that would have rivaled modern cetaceans in size (see the Perspective by Delsett and Pyenson). The animal existed at most 8 million years after the emergence of the first ichthyosaurs, suggesting a much more rapid size expansion that may have been fueled by processes after the Permian mass extinction. —SNV

Science, abf5787, this issue p. 1578; see also abm3751, p. 1554

ULTRAFAST MAGNETISM

Coupling up of spins and lattice

The development of spintronics and magnetic data storage relies on understanding and controlling the dynamics of magnetic excitations within a material. Of crucial importance for practical applications is how fast the magnetization can be switched. Mashkovich *et al*. report the use of ultrafast terahertz radiation to create magnon excitations in the antiferromagnet cobalt difluoride that can then be coupled with phonon excitations (see the Perspective by Juraschek and Narang). Using light to control coupling between the spins and the lattice provides a route to manipulating magnetization in antiferromagnetic materials on ultrafast time scales. —ISO

Science, abk1121, this issue p. 1608; see also abm0085, p. 1555

BIOCATALYSIS

Radical cyclization made easy

Metalloenzymes contain metal cofactors that can sometimes be exploited to catalyze reactions distinct from their natural function. Drawing on inspiration from heme-based radical polymerizations, Zhou *et al*. screened a panel of metalloproteins for candidates that could catalyze an atom-transfer radical cyclization, producing a lactam from an amide substrate bearing bromoalkyl and alkene functional groups (see the Perspective by Zhang and Dydio). Cytochrome P-450 derivatives were effective catalysts, and a series of enzyme variants were generated by directed evolution to access the full range of enantiomer and diastereomer products for model substrates. Such biocatalysts are valuable additions to the synthetic chemistry toolkit and reiterate the potential of metalloenzymes in catalyzing useful unnatural reactions. —MAF

Science, abk1603, this issue p. 1612; see also abm8321, p. 1558

CORONAVIRUS

How the Delta variant evades defenses

In the course of the COVID-19 epidemic, variants of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) continue to emerge, some of which evade immunity or increase transmission. In late 2020, the Delta and Kappa variants were detected, and the Delta variant became globally dominant by June 2021. McCallum *et al*. show that vaccine-elicited serum-neutralizing activity is reduced against these variants. Based on biochemistry and structural studies, the authors show that mutations in the domain that binds the ACE2 receptor abrogate binding to some monoclonal antibodies but do not improve ACE2 binding,

suggesting that they emerged to escape immune recognition. Remodeling of the N-terminal domain allows the variants to escape recognition by most neutralizing antibodies that target it. The work could guide the development of next-generation vaccines and antibody therapies. —VV

Science, abl8506, this issue p. 1621

PEROVSKITE STRUCTURE

Stable but not quite cubic

The black, photoactive phase of formamidinium (FA) perovskites, which is usually stabilized by cation alloying to avoid the formation of inactive hexagonal phases, is assumed to be cubic. High-resolution microscopy studies by Doherty *et al*. using nanoscale probes revealed that these FA-rich phases are not cubic but rather undergo slight tilting (by two degrees) of the octahedra. Black phases can have localized regions of hexagonal phases that nucleate degradation. Surface-bound ethylenediaminetetraacetic acid stabilized the tilted phase of pure FA lead triiodide against environmental degradation. —PDS

Science, abl4890, this issue p. 1598

CORONAVIRUS

A tool to probe SARS-CoV-2 biology

To develop therapies against severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and emerging variants, it is important to understand the viral biology and the effect of mutations. However, this is challenging because live virus can only be studied in a few laboratories that meet stringent safety standards. Syed *et al*. describe a virus-like particle (VLP) that comprises the four SARS-CoV-2 structural proteins, but instead of packaging viral RNA, it packages messenger RNA (mRNA) that expresses a reporter protein (see the Perspective by Johnson and Menachery). The amount of reporter expressed in receiver cells depends on the efficiency of packaging and assembly in the producer cells and the efficiency of entry into receiver cells. Mutations in the nucleocapsid protein that are found in more transmissible variants increase mRNA packaging and expression. The VLPs provide a platform for studying the effect of mutations in the structural proteins and for screening therapeutics. —VV

Science, abl6184, this issue p. 1626; see also abn3781, p. 1557

MUCOSAL IMMUNOLOGY

Stromal cells in training immunity

Immune responses to pathogens are heightened in individuals previously exposed to pathogen-associated stimuli, a phenomenon known as innate immune training. This process is accomplished through metabolic reprogramming and epigenetic changes, but the underlying mechanisms are not fully understood. Wu *et al.* investigated the functional requirement of interleukin (IL)-17 receptor signaling in fibroblastic reticular cells (FRCs) in the development of protective antibody responses against gut infection and studied how innate immune training affects these responses. Mild gut inflammatory training before bacterial infection enhanced gut-protective antibody production and reduced colon inflammation. Inflammatory training required IL-17 signaling in FRCs and IL-10 production by B cells. These findings reveal a mechanism of innate immune training in FRCs for protection against gut infection. —HMI

Sci. Immunol. **6**, eaao3669 (2021).

RESEARCH ARTICLE SUMMARY

IMMUNOLOGY

Fibrin is a critical regulator of neutrophil effector function at the oral mucosal barrier

Lakmali M. Silva, Andrew D. Doyle, Teresa Greenwell-Wild, Nicolas Dutzan, Collin L. Tran, Loreto Abusleme, Lih Jiin Juang, Jerry Leung, Elizabeth M. Chun, Andrew G. Lum, Cary S. Agler, Carlos E. Zuazo, Megan Sibree, Priyam Jani, Vardit Kram, Daniel Martin, Kevin Moss, Michail S. Lionakis, Francis J. Castellino, Christian J. Kastrup, Matthew J. Flick, Kimon Divaris, Thomas H. Bugge*, Niki M. Moutsopoulos*

INTRODUCTION: Mucosal barrier sites are continuously exposed to diverse environmental stimuli, and the fine-tuning of mucosal tissue–specific immune responses is key to maintaining tissue homeostasis. However, specific mechanisms of mucosal barrier immunity remain to be elucidated. In humans, congenital deficiency in plasminogen—the proenzyme form of the abundant plasma protease plasmin—leads to severe mucosal inflammatory disease with oral (ligneous periodontitis), ocular (ligneous conjunctivitis), lung, vaginal, and gastrointestinal tract involvement, which implicates plasmin in mucosal barrier homeostasis.

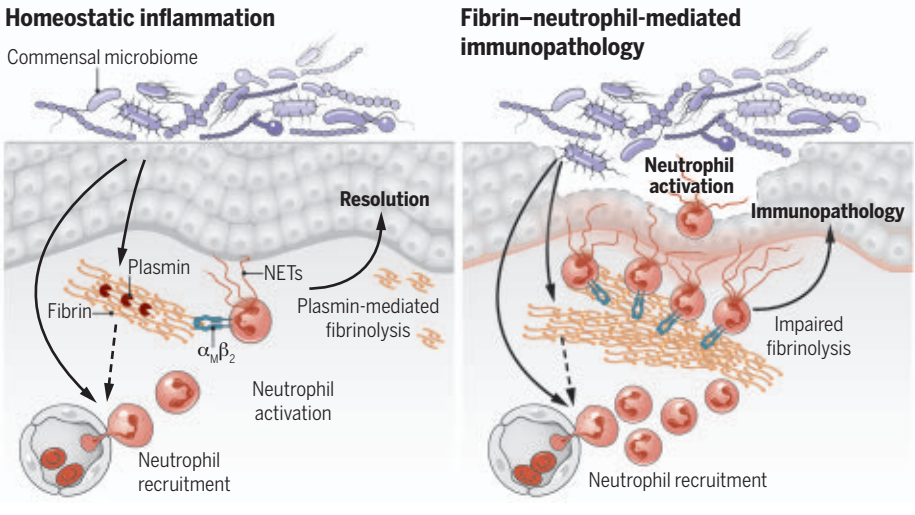
RATIONALE: Fibrin is the principal physiological target of plasmin, and abnormal inflammatory responses associated with plasminogen deficiency in humans and in animal models have been linked to extravascular fibrin deposition. In the oral cavity, excess extravascular fibrin is hypothesized to underlie severe oral mucosal (gingival) inflammation, supporting bone (periodontal) destruction, and adolescent loss of dentition associated with plasminogen deficiency in humans. Mucosal inflammation surrounding dentition and destruction of underlying bone are also hallmarks of common forms of periodontitis. In this study, we sought to determine the mechanistic link between fibrin deposition and oral mucosal immunopathology.

RESULTS: We employed an array of genetically engineered mouse models and pharmacological inhibitors, complemented by histological and genetic studies in humans, to probe the association between fibrin deposition and local immunopathology. Consistent with findings in plasminogen-deficient humans, mice lacking plasminogen displayed extravascular fibrin accumulation, gingival inflammation, and periodontal bone loss. Oral immunopathology in plasminogen deficiency was prevented by genetic or pharmacological elimination of the fibrin precursor, fibrinogen, demonstrating the critical role of fibrin in plasminogen deficiency–associated mucosal immunopathology.

Notably, oral immunopathology was significantly reduced in plasminogen-deficient germ-free mice compared with specific pathogen–free mice. This suggests that commensal microbial communities represent a trigger for oral mucosal fibrin deposition and associated immunopathology.

Oral lesions in plasminogen-deficient mice were sites of neutrophil accumulation proximal to fibrin deposits. Notably, plasminogen-deficient mice expressing a mutant fibrin lacking the binding site for the myeloid integrin $\alpha_M\beta_2$ did not develop periodontal disease, despite featuring unabated gingival fibrin deposition and neutrophil accumulation. Fibrin engagement by $\alpha_M\beta_2$ promoted neutrophil effector functions, including the production of reactive oxygen species and neutrophil extracellular trap (NET) formation (NETosis). Accordingly, genetic or pharmacological ablation of NETosis reduced periodontal bone destruction in plasminogen-deficient mice. Thus, neutrophil activation, through fibrin-neutrophil engagement, triggered oral mucosal immunopathology. Abrogation of fibrin-neutrophil engagement also reduced age-associated periodontal immunopathology under normal fibrinolytic conditions, which was suggestive of a broader role for this mechanism in non-Mendelian forms of periodontitis. Consistent with this notion, variation in the human *PLG* gene, which encodes plasminogen, was associated with increased prevalence of common forms of severe periodontal disease in a cohort of Americans of European ancestry.

CONCLUSION: We identify fibrin as a critical immune regulator of oral mucosal barrier homeostasis, mediated by the local engagement and activation of neutrophils. These findings provide insights into the role of fibrin in neutrophil activation and oral mucosal immunity under both defective and normal fibrinolysis. Our findings pave the way for the investigation of fibrin-neutrophil engagement as a target for the prevention or treatment of the prevalent human disease periodontitis. ■



Fibrin-neutrophil engagement mediates oral mucosal immunopathology. In the oral mucosa, commensal microbiota triggers homeostatic inflammation, fibrin deposition, and associated neutrophil activation, which physiologically resolves in the setting of efficient plasminogen-mediated fibrinolysis (left). When fibrinolysis is impaired, fibrin accumulation leads to persistent activation of neutrophils and promotes local immunopathology (right). Dashed arrows denote mechanisms that are not yet fully defined.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: thomas.bugge@nih.gov (T.H.B.); nmoutsop@mail.nih.gov (N.M.M.)
Cite this article as L. M. Silva *et al.*, *Science* **374**, eabl5450 (2021). DOI: 10.1126/science.abl5450

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abl5450

RESEARCH ARTICLE

IMMUNOLOGY

Fibrin is a critical regulator of neutrophil effector function at the oral mucosal barrier

Lakmali M. Silva[1,2], Andrew D. Doyle[3], Teresa Greenwell-Wild[2], Nicolas Dutzan[2,4], Collin L. Tran[1], Loreto Abusleme[2,5], Lih Jiin Juang[6], Jerry Leung[6], Elizabeth M. Chun[1], Andrew G. Lum[2], Cary S. Agler[7], Carlos E. Zuazo[2]†, Megan Sibree[1], Priyam Jani[8], Vardit Kram[8], Daniel Martin[9], Kevin Moss[7], Michail S. Lionakis[10], Francis J. Castellino[11], Christian J. Kastrup[6,12,13], Matthew J. Flick[14], Kimon Divaris[15,16], Thomas H. Bugge[1]*, Niki M. Moutsopoulos[2]*

Tissue-specific cues are critical for homeostasis at mucosal barriers. Here, we report that the clotting factor fibrin is a critical regulator of neutrophil function at the oral mucosal barrier. We demonstrate that commensal microbiota trigger extravascular fibrin deposition in the oral mucosa. Fibrin engages neutrophils through the $\alpha_M\beta_2$ integrin receptor and activates effector functions, including the production of reactive oxygen species and neutrophil extracellular trap formation. These immune-protective neutrophil functions become tissue damaging in the context of impaired plasmin-mediated fibrinolysis in mice and humans. Concordantly, genetic polymorphisms in *PLG*, encoding plasminogen, are associated with common forms of periodontal disease. Thus, fibrin is a critical regulator of neutrophil effector function, and fibrin-neutrophil engagement may be a pathogenic instigator for a prevalent mucosal disease.

The study of Mendelian diseases is instrumental in uncovering genes and pathways that are implicated in human immunity and mucosal homeostasis. One Mendelian disease clearly linked to mucosal immunopathology is plasminogen (PLG) deficiency. PLG is synthesized by the liver and is present at high concentrations in the circulation and in interstitial fluids. PLG is a zymogen, which is converted to the active serine protease plasmin by urokinase or tissue plasminogen activator. Homozygous or compound heterozygous mutations in the human *PLG* gene typically lead to severe mucosal disease, suggesting a critical role for this gene in mucosal immunity. Such patients present with deposition of fibrin at various mucosal sites leading to oral disease (ligneous periodontitis); ocular disease (conjunctivitis); and lung, vaginal, and gastrointestinal disease (*1*–*3*). Fibrin, a physiological target of plasmin, is by itself a proinflammatory molecule, and many of the abnormal inflammatory responses associated with PLG deficiency have been linked to persistent extravascular fibrin deposition. In the oral mucosa, local deposition of fibrin is hypothesized to lead to severe soft tissue and bone destruction around teeth and often loss of the entire dentition in adolescence (*4*–*7*). Mucosal inflammation in areas surrounding the dentition and destruction of underlying bone are also the hallmarks of the common human oral mucosal disease, periodontitis (*8*, *9*).

We hypothesized that insufficient plasmin-mediated fibrinolysis is a key contributor to mucosal immunopathology in humans and in animal models. To understand the mechanistic links between mucosal fibrin deposition and immunopathology, we employed an array of genetically engineered mouse models, complemented by histological and genetic studies in humans. We chose to study the mechanisms involved in PLG deficiency–mediated immunopathology at the site of the oral mucosa because periodontitis is among the most common manifestations in humans with PLG deficiency. This suggests that this mucosal site is particularly susceptible to PLG deficiency–related immunopathology. Our work highlights fibrin as a critical immune regulator of mucosal barrier homeostasis and provides insights into the role of fibrin in neutrophil activation and tissue immunity.

Results

Plg *deficiency instigates oral mucosal immunopathology in mice*

$Plg^{-/-}$ mice phenocopy PLG-deficient humans and present with mucosal immunopathology, including periodontitis, conjunctivitis, and gastrointestinal immunopathology (*10*–*13*) (Fig. 1A and fig. S1). The hallmark of periodontitis is mucosal inflammation and destruction of bone around the teeth. $Plg^{-/-}$ mice developed periodontal bone loss by 12 weeks of age compared with their wild-type and heterozygous littermates, and this loss became increasingly severe at 24 weeks of age (Fig. 1, A and B, and fig. S2A). Consistent with accelerated periodontal bone loss in $Plg^{-/-}$ mice, mice double-deficient for the two physiological PLG activators, tissue plasminogen activator and urokinase plasminogen activator ($Plat^{-/-}$;$Plau^{-/-}$), also displayed severe periodontal bone loss (fig. S2B). To further confirm that spontaneous periodontitis in $Plg^{-/-}$ mice was associated with the loss of enzymatic activity of plasmin, we used a knock-in mouse model that harbors a mutation in the catalytic serine residue (S743A), which renders plasmin catalytically inactive ($Plg^{S743A/S743A}$). At the ages of 12 and 24 weeks, $Plg^{S743A/S743A}$ mice showed significant increases in periodontal bone loss compared with their wild-type and heterozygous littermates, confirming that plasmin insufficiency drives periodontal bone loss (Fig. 1C). Consistent with mucosal lesions in patients with PLG deficiency, the oral mucosal tissues of $Plg^{-/-}$ mice displayed a disruption of tissue architecture associated with the deposition of fibrin (Fig. 1, D and E). Beyond fibrin deposition, the oral mucosal immunopathology in $Plg^{-/-}$ mice was characterized by mucosal edema and loss of connective tissue attachment with disruption of the junctional epithelium (fig. S3A). Thus, we confirm that *Plg* deficiency in mice, like human disease, leads to severe oral mucosal immunopathology associated with fibrin deposition.

Elimination of fibrinogen rescues oral mucosal immunopathology in Plg*-deficient mice*

To determine whether fibrin accumulation, secondary to impaired plasmin-mediated fibrinolysis, was the trigger for mucosal immunopathology, we generated *Plg* and fibrinogen

[1]Proteases and Tissue Remodeling Section, National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, MD, USA. [2]Oral Immunity and Inflammation Section, National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, MD, USA. [3]NIDCR Imaging Core, National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, MD, USA. [4]Department of Conservative Dentistry, Faculty of Dentistry, University of Chile, Santiago, Chile. [5]Department of Pathology and Oral Medicine, Faculty of Dentistry, University of Chile, Santiago, Chile. [6]Michael Smith Laboratories and Department of Biochemistry and Molecular Biology, The University of British Columbia, Vancouver, BC, Canada. [7]Division of Oral and Craniofacial Health Sciences, Adams School of Dentistry, University of North Carolina, Chapel Hill, NC, USA. [8]Molecular Biology of Bones and Teeth Section, National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, MD, USA. [9]NIDCR Genomics and Computational Biology Core, National Institute of Dental and Craniofacial Research, National Institutes of Health, Bethesda, MD, USA. [10]Fungal Pathogenesis Section, Laboratory of Clinical Immunology and Microbiology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD, USA. [11]WM Keck Center for Transgene Research, University of Notre Dame, Notre Dame, IN, USA. [12]Blood Research Institute, Versiti, Milwaukee, WI, USA. [13]Departments of Surgery, Biochemistry, Biomedical Engineering, and Pharmacology and Toxicology, Medical College of Wisconsin, Milwaukee, WI, USA. [14]Department of Pathology and Laboratory Medicine, University of North Carolina, Chapel Hill, NC, USA. [15]Division of Pediatric and Public Health, Adams School of Dentistry, University of North Carolina, Chapel Hill, NC, USA. [16]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, Chapel Hill, NC, USA.
*Corresponding author. Email: thomas.bugge@nih.gov (T.H.B.); nmoutsop@mail.nih.gov (N.M.M.)
†Present address: Department of Cell and Tissue Biology, University of California, San Francisco, CA, USA.

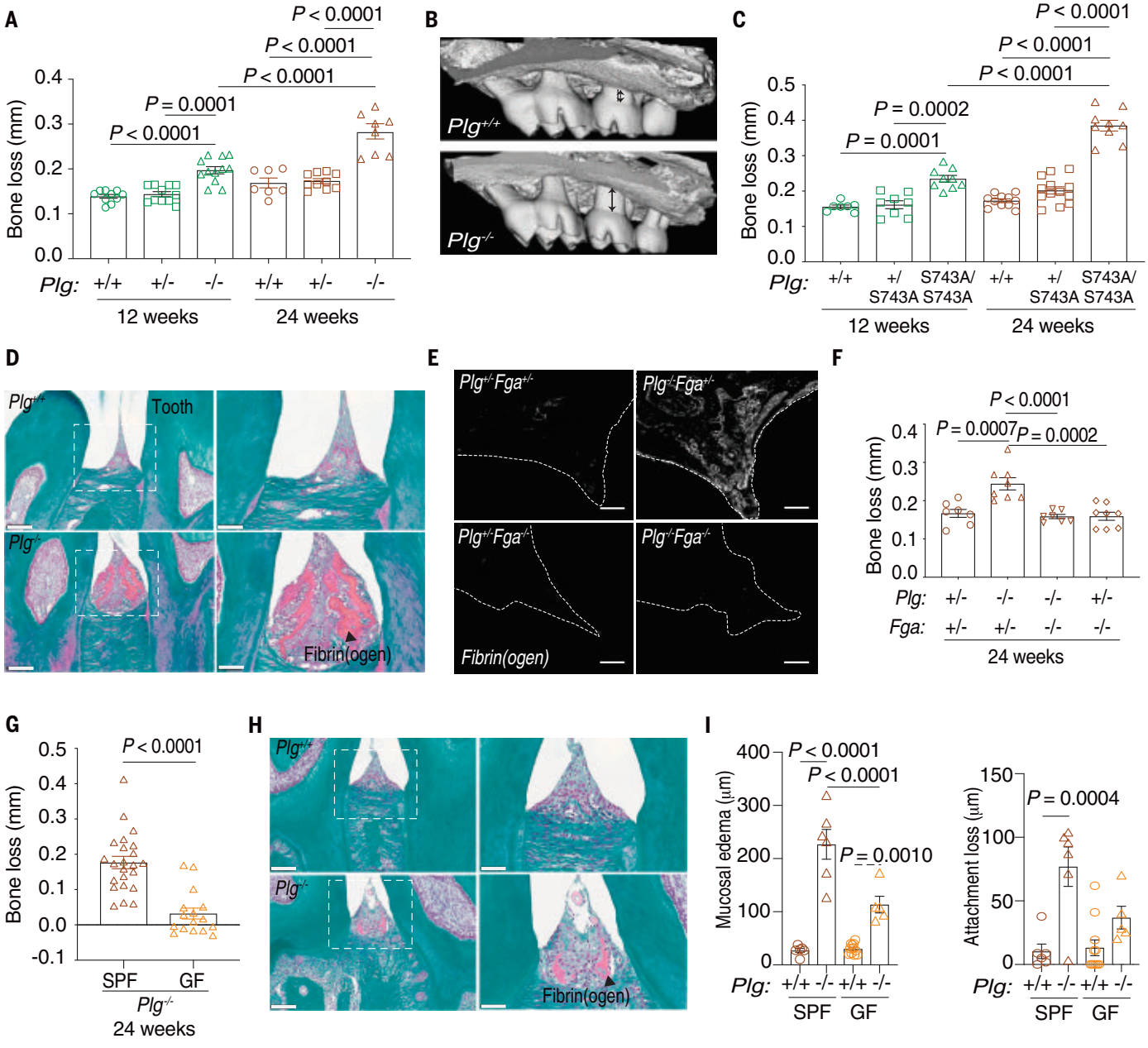


Fig. 1. Defective fibrinolysis triggers oral mucosal immunopathology. (**A**) *Plg*-deficient mice ($Plg^{-/-}$) display periodontal bone loss at the ages of 12 and 24 weeks compared with $Plg^{+/+}$ and $Plg^{+/-}$ littermates. (**B**) Micro-CT visualization of periodontal bone loss in 24-week-old $Plg^{+/+}$ and $Plg^{-/-}$ maxillae (black arrows depict the distance between alveolar bone crest and cementoenamel junction). (**C**) Mice with catalytically inactive plasmin ($Plg^{S743A/S743A}$) develop significantly increased spontaneous periodontal bone loss. (**D**) Fraser-Lendrum staining of $Plg^{+/+}$ and $Plg^{-/-}$ oral mucosal tissue sections (green indicates collagen, magenta indicates keratin and fibrin, orange indicates erythrocytes, and the black arrowhead depicts fibrin deposition). Scale bars, 100 μm (left) and 50 μm (right, insets). (**E**) Immunofluorescence (IF) staining for fibrin(ogen) (gray) in $Plg^{+/-}$;$Fga^{+/-}$, $Plg^{-/-}$;$Fga^{+/-}$, $Plg^{+/-}$;$Fga^{-/-}$, and $Plg^{-/-}$;$Fga^{-/-}$ oral tissue sections. Scale bars, 50 μm. (**F**) Bone loss measurements in $Plg^{+/-}$;$Fga^{+/-}$, $Plg^{-/-}$;$Fga^{+/-}$, $Plg^{+/-}$;$Fga^{-/-}$, and $Plg^{-/-}$;$Fga^{-/-}$ mice maxillae. (**G**) Difference (Δ) in bone loss between $Plg^{-/-}$ and $Plg^{+/+}$ mice under SPF and GF conditions. (**H**) Fraser-Lendrum staining of $Plg^{+/+}$ and $Plg^{-/-}$ GF mice oral mucosal tissue sections. Scale bars, 100 μm (left) and 50 μm (right, insets). (**I**) Measurements of mucosal edema and attachment loss in $Plg^{+/+}$ and $Plg^{-/-}$ under SPF and GF conditions.

double-deficient ($Plg^{-/-}$;$Fga^{-/-}$) mice. As expected, $Plg^{-/-}$;$Fga^{-/-}$ mice did not exhibit fibrin deposition in mucosal tissues, in contrast to their $Plg^{-/-}$;$Fga^{+/-}$ littermates (Fig. 1E). Notably, fibrinogen deficiency completely prevented the periodontal bone loss imposed by *Plg* deficiency (Fig. 1F). Thus, oral mucosal immunopathology in *Plg* deficiency is driven by fibrin accumulation.

The oral microbiome instigates fibrin-mediated immunopathology in Plg*-deficient mice*

To assess the contribution of the microbiome in fibrin-mediated mucosal immunopathology, we rederived $Plg^{-/-}$ mice in a germ-free (GF) setting. GF $Plg^{-/-}$ mice displayed significantly reduced mucosal fibrin deposition and associated mucosal immunopathology and periodontal bone loss compared with their specific pathogen–free (SPF) counterparts (Fig. 1, G to I, and fig. S3), which suggests a critical role for

the commensal microbiome. Consistent with these findings, antibiotic treatment of mice with defects in fibrinolysis (*Plg*$^{-/-}$, *Plat*$^{-/-}$; *Plau*$^{-/-}$, and *Plg*$^{S743A/S743A}$) also significantly ameliorated disease (fig. S4, A to C). *Plg*$^{-/-}$ mice did not exhibit any significant shifts in oral microbiome community structure and composition, nor any specific microbial constituents associated with triggering immunopathology (at 12 and 16 weeks of age; fig. S4, D to J), which is consistent with a lack of mucosal bacterial infection. Thus, the commensal microbiome is a trigger for oral mucosal immunopathology in the setting of defective fibrinolysis.

Oral mucosal lesions in Plg*-deficient mice are neutrophil dominated*

We next aimed to dissect the mechanisms of fibrin-mediated mucosal immunopathology. To this end, we took an unbiased approach and evaluated global transcriptomic changes by RNA sequencing (RNA-seq) of oral mucosal (gingival) tissues of *Plg*$^{-/-}$ mice compared with wild-type littermates. Principal components analysis of RNA-seq data revealed a distinct transcriptome in the mucosal tissues of *Plg*$^{-/-}$ mice (Fig. 2A, fig. S5, and table S1). Gene ontology (GO) analysis of the most significantly up-regulated pathways in *Plg* deficiency revealed neutrophil migration as the top differentially expressed pathway, followed by inflammation, extracellular matrix disassembly, cytokine signaling, leukocyte migration, and response to bacterial molecules (Fig. 2B and table S2). Examination of the most significantly differentially expressed genes between *Plg*$^{-/-}$ and wild-type mice revealed up-regulation of genes associated with neutrophil recruitment and granulopoiesis, including chemokine (C-X-C motif) ligand (*Cxcl*)*1* to *3* and *5*, colony-stimulating factor 3 (*Csf3*), and Csf3 receptor (*Csf3r*). Genes involved in neutrophil and monocyte activation were also up-regulated, including pro-platelet basic protein (*Ppbp*), ADAM metallopeptidase domain 8 (*Adam8*), triggering receptor expressed on myeloid cells 1 (*Trem1*), C-type lectin domain family 4 member E (*Clec4e*), and CD300 molecule like family member B (*Cd300lb*) (Fig. 2C). Finally, several matrix metalloproteases (*Mmp8* to *10* and *12*), interleukins (*Il1a* and *Il1b*), and prostaglandin-endoperoxide synthase 2 (*Ptgs2*) that are involved in extracellular matrix remodeling and inflammation were also up-regulated in *Plg*$^{-/-}$ mice.

We next evaluated the gingival inflammatory infiltrate by flow cytometry (fig. S6A). Consistent with the transcriptomic findings, the oral mucosal inflammatory infiltrate in *Plg*$^{-/-}$ mice was dominated by neutrophils (live CD45^{+}CD11b^{+}Cd11c$^{neg\text{-}mid}$Ly6G^{+}) at both early (12 weeks) and late (24 weeks) time points (Fig. 2, D and E). The absolute number of neutrophils was significantly increased in *Plg*$^{-/-}$ mice at all time points, whereas their proportion increased significantly in late stages of immunopathology. Notably, neutrophils were the only cell subset that significantly increased at early time points of immunopathology (12 weeks).

There was no increase in monocyte or macrophage cells (live SSC^{int}CD45^{+}CD11b^{+}Cd11c$^{neg\text{-}mid}$ Ly6G^{-}) at 12 weeks, although absolute counts increased at late stages of immunopathology (24 weeks) (Fig. 2, D and F). B cell (live CD45^{+}B220^{+}), T cell (live CD45^{+}TCRβ^{+}), and dendritic cell (live CD45^{+}CD11C^{hi}) numbers and proportions did not change at any stage of disease (fig. S6, B to D). Immunostaining for fibrin(ogen) and Ly6G—a marker for neutrophils—in gingival tissues confirmed that neutrophils accumulated in increased numbers in gingival tissues of *Plg*$^{-/-}$ mice and were often found in proximity to fibrin(ogen) deposits (Fig. 2G). Thus, oral mucosal lesions in *Plg*$^{-/-}$ mice are dominated by neutrophils and likely reflect a neutrophil-mediated immunopathology.

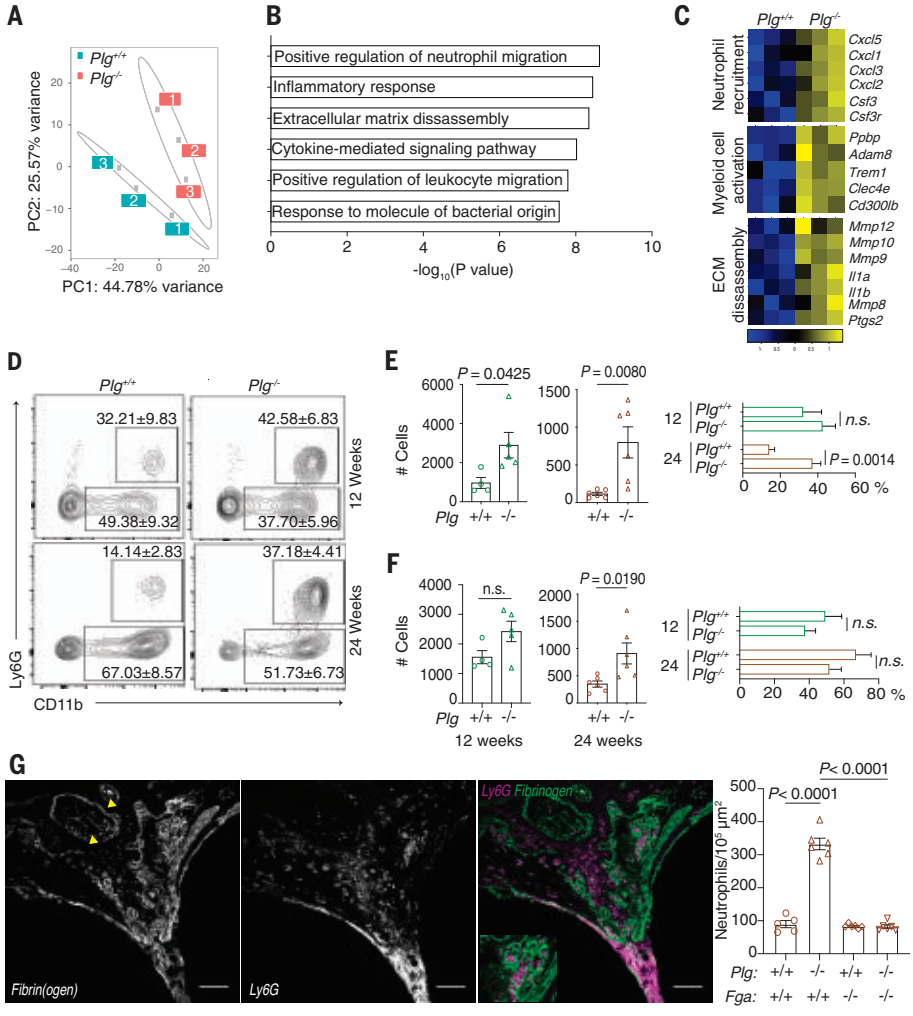


Fig. 2. Lesions of mucosal immunopathology in *Plg* deficiency are dominated by neutrophils. (**A**) Principal coordinates analysis (PCoA) plot showing the distribution of differentially expressed genes in oral mucosa of 12-week-old *Plg*$^{+/+}$ and *Plg*$^{-/-}$ mice. (**B**) GO biological processes up-regulated in *Plg*$^{-/-}$ mice in ascending order of *P* value. (**C**) Heatmaps showing significantly differentially expressed genes involved in granulopoiesis or neutrophil recruitment, myeloid cell activation, and inflammation or extracellular matrix (ECM) disassembly in *Plg*$^{-/-}$ compared with *Plg*$^{+/+}$ gingival tissues. (**D**) Flow cytometry analysis of 12- and 24-week-old mouse oral mucosal tissues. Contour plots show changes in neutrophil and monocyte or macrophage populations in percentages. (**E** and **F**) Absolute count and percentages of neutrophils (E) and monocytes or macrophages (F) in 12- and 24-week-old *Plg*$^{+/+}$ and *Plg*$^{-/-}$ mouse gingiva. (**G**) Fibrin(ogen) (green) and Ly6G (neutrophils; magenta) IF staining of 24-week-old *Plg*$^{-/-}$ mouse oral mucosal sections showing extravascular localization of fibrin deposits and neutrophils. Yellow arrowheads indicate blood vessels. Scale bars, 50 µm. On the right, there is a quantification of the number of neutrophils per 10^5 µm^2 of stained tissue sections.

Fibrin(ogen)-mediated periodontal immunopathology depends on its myeloid integrin $\alpha_M\beta_2$-binding motif

We next aimed to test whether neutrophil engagement of fibrin through $\alpha_M\beta_2$ in gingival tissues is critical for neutrophil retention and/or activation and associated immunopathology. Periodontal bone loss was completely prevented in *Plg*$^{-/-}$ mice by the elimination of the $\alpha_M\beta_2$-binding motif from fibrin(ogen) (*Plg*$^{-/-}$;*Fgg*$^{390\text{-}396A/390\text{-}396A}$) compared with *Plg*$^{-/-}$ and *Plg*$^{-/-}$;*Fgg*$^{+/390\text{-}396A}$ littermates (Fig. 3, A and B). The prevention of periodontal bone loss in *Plg*$^{-/-}$;*Fgg*$^{390\text{-}396A/390\text{-}396A}$ mice was not a consequence of diminished gingival fibrin accumulation, as gingival tissues showed increased fibrin(ogen) accumulation in *Plg*$^{-/-}$;*Fgg*$^{+/390\text{-}396A}$ and *Plg*$^{-/-}$;*Fgg*$^{390\text{-}396A/390\text{-}396A}$ mice compared with *Plg*$^{+/-}$;*Fgg*$^{+/390\text{-}396A}$ littermates (Fig. 3, C and D). Prevention of alveolar bone loss in *Plg*$^{-/-}$;*Fgg*$^{+/390\text{-}396A}$ mice was also not a consequence of abolished gingival neutrophil accumulation, as tissue immunostaining (Ly6G) (Fig. 3, C and E) and flow cytometry (Fig. 3, F and G) demonstrated that lesions of *Plg*$^{-/-}$;*Fgg*$^{390\text{-}396A/390\text{-}396A}$ mice accumulated neutrophils in a manner comparable to those of *Plg*$^{-/-}$;*Fgg*$^{+/390\text{-}396A}$ littermates.

Plg$^{+/-}$;*Fgg*$^{390\text{-}396A/390\text{-}396A}$ mice showed a reduction of bone loss compared with their *Plg*$^{+/-}$;*Fgg*$^{+/390\text{-}396A}$ littermates (Fig. 3A). Thus, we examined whether the fibrin(ogen)–$\alpha_M\beta_2$ integrin interaction contributes to periodontal bone loss also under normal fibrinolytic conditions. Wild-type mice are known to develop spontaneous periodontal bone loss with age, which becomes significant between 10 and 24 weeks (Fig. 3H) (*14*). However, *Fgg*$^{390\text{-}396A/390\text{-}396A}$ mice showed significantly less periodontal bone loss than *Fgg*$^{+/+}$ and *Fgg*$^{+/390\text{-}396A}$ littermate controls (Fig. 3I). Thus, the local engagement of fibrin(ogen) by neutrophils through $\alpha_M\beta_2$ is essential for mucosal immunopathology and associated periodontal bone loss under defective or normal fibrinolytic conditions.

Engagement of fibrin(ogen) through $\alpha_M\beta_2$ integrin regulates neutrophil effector functions

Next, we established an in vitro system to study the effects of engagement of neutrophils with fibrin(ogen) purified from either wild-type or *Fgg*$^{390\text{-}396A/390\text{-}396A}$ mice. In agreement with previous studies (*15*), human and mouse neutrophils showed a reduced adherence to mutant fibrin compared with wild-type fibrin (Fig. 4A and fig. S7A). A similar reduction in adherence was noted when α_M-deficient neutrophils

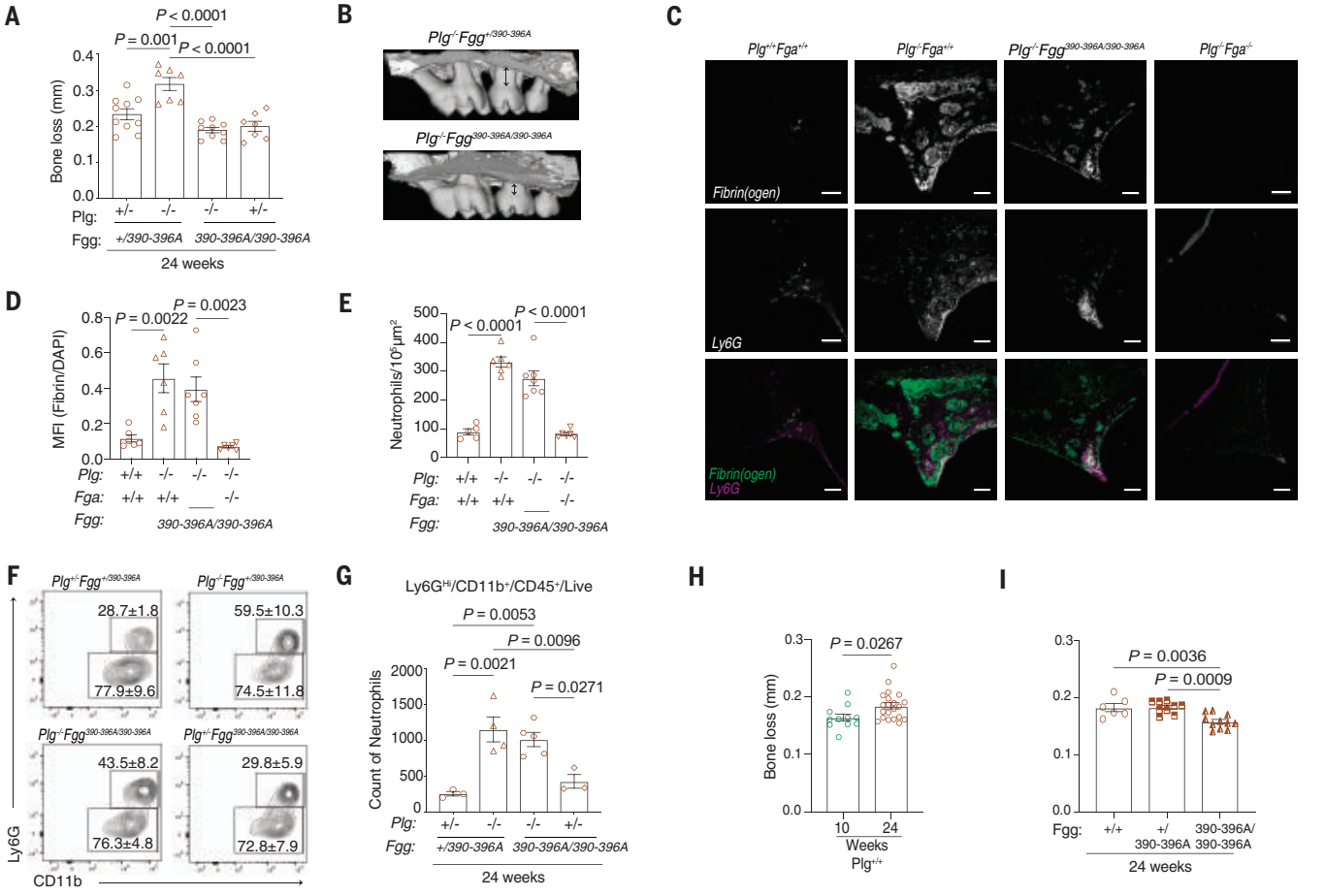


Fig. 3. Myeloid integrin $\alpha_M\beta_2$-binding motif on fibrin(ogen) mediates plasminogen deficiency–induced oral immunopathology. (**A**) Bone loss measurement in *Plg*$^{+/-}$;*Fgg*$^{+/390\text{-}396A}$, *Plg*$^{-/-}$;*Fgg*$^{+/390\text{-}396A}$, *Plg*$^{+/-}$;*Fgg*$^{390\text{-}396A/390\text{-}396A}$, and *Plg*$^{-/-}$;*Fgg*$^{390\text{-}396A/390\text{-}396A}$ mice maxillae. (**B**) Micro-CT visualization of periodontal bone loss of maxilla (black arrows depict the distance between alveolar bone crest and cementoenamel junction). (**C**) IF staining of fibrin(ogen) (green) and Ly6G (neutrophils; magenta) in *Plg*$^{+/+}$;*Fga*$^{+/+}$, *Plg*$^{-/-}$;*Fga*$^{+/+}$, *Plg*$^{-/-}$;*Fgg*$^{390\text{-}396A/390\text{-}396A}$, and *Plg*$^{-/-}$;*Fga*$^{-/-}$ mouse gingival sections. Scale bars, 50 µm. (**D** and **E**) Graphs showing the mean fluorescence intensity (MFI) of fibrin(ogen) or DAPI staining (D) and the number of neutrophils per 10^5 µm^2 of stained tissue sections (E). (**F**) Flow cytometry analysis of 24-week-old gingiva. Contour plots show neutrophil and monocyte or macrophage populations in percentages. (**G**) Counts of neutrophils in 24-week-old *Plg*$^{+/-}$;*Fgg*$^{+/390\text{-}396A}$, *Plg*$^{-/-}$;*Fgg*$^{+/390\text{-}396A}$, *Plg*$^{-/-}$;*Fgg*$^{390\text{-}396A/390\text{-}396A}$, and *Plg*$^{+/-}$;*Fgg*$^{390\text{-}396A/390\text{-}396A}$ mouse gingival sections. (**H**) Bone loss measurement in wild-type 10- and 24-week-old maxillae. (**I**) Bone loss measurement in *Fgg*$^{+/+}$, *Fgg*$^{+/390\text{-}396A}$ and *Fgg*$^{390\text{-}396A/390\text{-}396A}$ mice maxillae at 24 weeks of age.

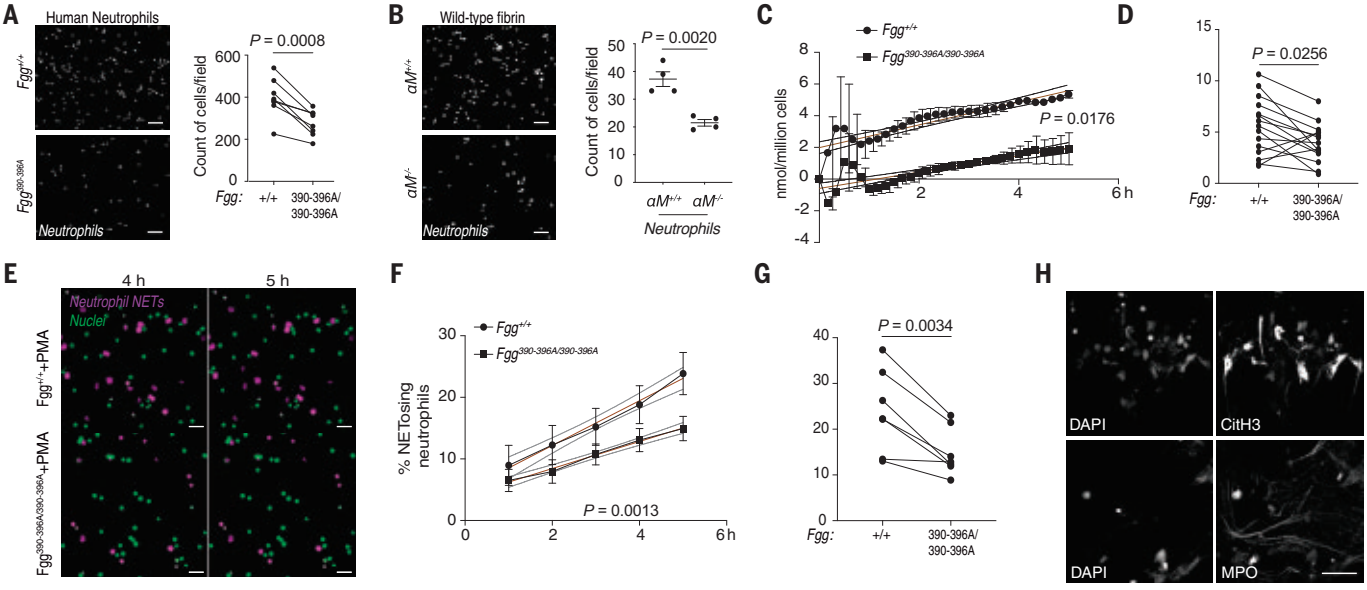
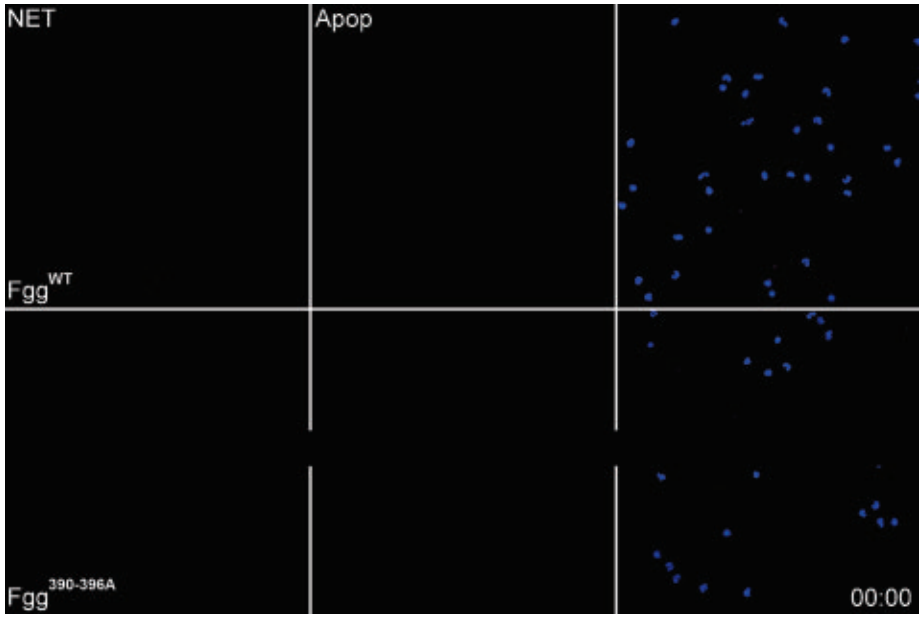


Fig. 4. Fibrin(ogen)-neutrophil interaction through $\alpha_M\beta_2$-binding triggers key neutrophil effector functions. (**A**) Human neutrophil binding to *Fgg*$^{390\text{-}396A/390\text{-}396A}$ fibrin compared with wild-type fibrin. Scale bars, 200 μm. (**B**) Binding of wild-type and α_M integrin–deficient ($\alpha_M^{-/-}$) mouse neutrophils to wild-type fibrin. Scale bars, 200 μm. (C to H) Neutrophils were plated on wild-type or Fgg$^{390\text{-}396A}$ fibrin. (**C**) Representative graph of neutrophil ROS production over time (15 nM MPO). (**D**) ROS production at 5-hour time point (n = 15). (**E**) NETs (magenta) at 4- and 5-hour time points for neutrophils plated on wild-type and Fgg$^{390\text{-}396A}$ fibrin. Scale bars, 50 μm. (**F**) NETosis progression over time (1 to 5 hours) (see Movie 1). (**G**) Quantification of NETosing cells at 5-hour time point (n = 7). (**H**) Cit-H3 and MPO staining (gray) of NETs at 5-hour time point. Scale bars, 50 μm.

were plated on wild-type fibrin and compared with wild-type neutrophils (Fig. 4B). Thus, $\alpha_M\beta_2$ is important for neutrophil adhesion to fibrin. To assess the effect of neutrophil engagement of fibrin on neutrophil effector functions, we studied the production of reactive oxygen species (ROS) and neutrophil extracellular trap (NET) formation (NETosis) by human neutrophils plated on wild-type or mutant fibrin. We found that when neutrophils were in association with wild-type fibrin, ROS production per cell (Fig. 4, C and D) was significantly increased compared with neutrophils plated on fibrin lacking the $\alpha_M\beta_2$-binding motif. The proportion of NETosis events, as visualized by an increase in cellular size and subsequent externalization of DNA (Fig. 4, E to G; fig. S7, B and C; and Movie 1), was also significantly increased when neutrophils plated on wild-type fibrin were compared with fibrin lacking the $\alpha_M\beta_2$-binding motif. Fibrin-induced NETosis was further confirmed by positive staining for extracellular citrullinated histone (Cit-H3) and myeloperoxidase (MPO) (Fig. 4H and fig. S7D). Consistent with fibrin-mediated NETosis in vivo, we observed enhanced diffuse extracellular staining for Cit-H3 and MPO in *Plg*$^{-/-}$ mice (Fig. 5A). Extracellular Cit-H3 and MPO were markedly reduced when the fibrin-neutrophil interaction was abrogated in *Plg*$^{-/-}$;*Fgg*$^{390\text{-}396A/390\text{-}396A}$ mice (Fig. 5A). Thus, the local engagement of fibrin (ogen) by neutrophils through $\alpha_M\beta_2$ is important for neutrophil activation.



Movie 1. NETosis with wild-type or Fgg$^{390\text{-}396A}$ fibrin. Random video for a time-lapse series showing live neutrophils (Hoechst; blue) and those undergoing either NETosis (magenta overlay) or apoptosis (cytox green; green) after stimulation with 15 nM PMA plated on either wild-type or Fgg$^{390\text{-}396A}$ fibrin.

Fibrin-induced NETosis mediates periodontal immunopathology in Plg*-deficient mice*

To understand whether NETosis contributes to the periodontal disease phenotype in vivo, we inhibited the accumulation of extracellular DNA and NETosis in *Plg*$^{-/-}$ mice with pharmacologic and genetic tools. *Plg*$^{-/-}$ mice treated with deoxyribonuclease (DNase) I intraperitoneally (i.p.) from weeks 8 to 20 exhibited reduced alveolar bone loss compared with the

control group (Fig. 5B). Removal (or reduction) of NETs by systemic DNase I treatment (i.p.) was evident in tissues of DNase I–treated *Plg*$^{-/-}$ mice by the significant reduction of extracellular Cit-H3 and MPO (Fig. 5C and fig. S8, A, B, and E). Next, we generated *Plg*-deficient mice lacking neutrophil elastase (*Plg*$^{-/-}$;*Ela*$^{-/-}$), which is required for NETosis (*16*). Deletion of *Ela* abrogated NETosis in vivo (fig. S8, C to D and F) and significantly reduced bone loss in *Plg*$^{-/-}$ mice (Fig. 5D). Thus, neutrophil NETosis in vivo is a mediator of fibrin-triggered oral mucosal immunopathology.

Finally, small interfering RNAs (siRNAs) targeted to deplete fibrinogen (siFbg), delivered through lipid nanoparticles (*17*) produced a reduction of circulating fibrinogen (Fig. 5E), which was associated with significant protection from periodontal bone loss (Fig. 5F). Thus, either inhibition of fibrin accumulation or NETosis were effective in preventing periodontal bone loss.

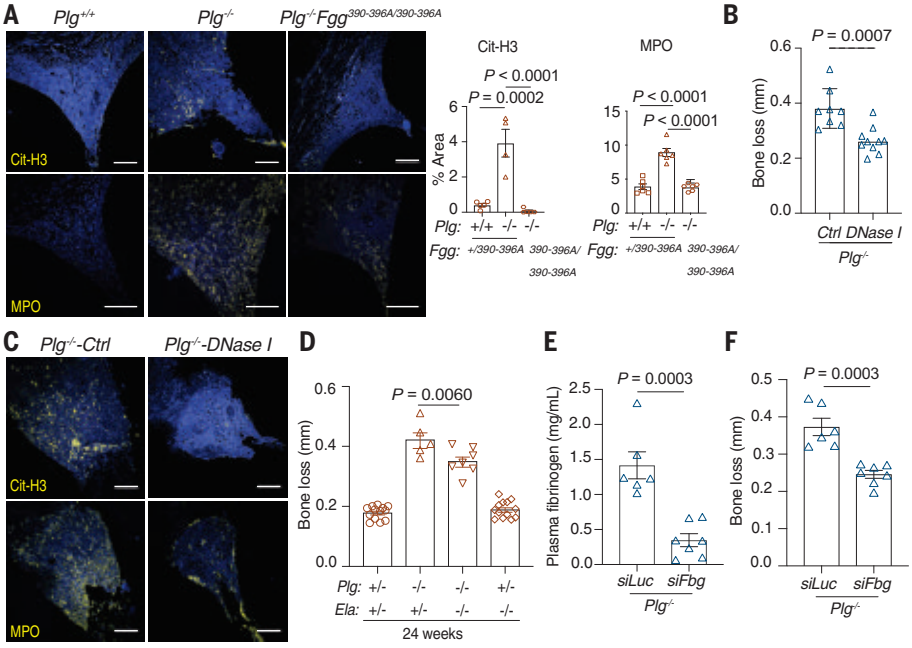


Fig. 5. NETosis mediates periodontal immunopathology in *Plg*-deficient mice. (**A**) (Left) Immunofluorescence staining of Cit-H3 and MPO (yellow) in *Plg*$^{+/+}$, *Plg*$^{-/-}$, and *Plg*$^{-/-}$;*Fgg*$^{390\text{-}396A/390\text{-}396A}$ mouse oral mucosal tissue sections. Scale bars, 50 μm. (Right) Quantification of percentage of Cit-H3 and MPO stained area, respectively. (**B** and **C**) *Plg*$^{-/-}$ control (Ctrl) and *Plg*$^{-/-}$ DNase I–treated mice. (B) Bone loss quantification. (C) Staining for Cit-H3 and MPO (yellow) in gingiva. Scale bars, 50 μm. (**D**) *Plg*$^{+/-}$;*Ela*$^{+/-}$, *Plg*$^{-/-}$;*Ela*$^{+/-}$, *Plg*$^{-/-}$;*Ela*$^{-/-}$, and *Plg*$^{+/-}$;*Ela*$^{-/-}$ mice bone loss measurements. (**E** and **F**) siRNA treatment. (E) Plasma fibrinogen levels. (F) Bone-loss measurements of *Plg*$^{-/-}$-siLuc– and *Plg*$^{-/-}$-siFbg–treated mice.

PLG *polymorphisms are associated with the common oral mucosal disease periodontitis*

In periodontitis, local tissue inflammation and bleeding are hallmarks of disease (*18*). Additionally, the infiltration of neutrophils is a characteristic of this condition (*19*). Consistent with the notion that neutrophil-fibrin interactions could also mediate periodontitis in patients, we found accumulation of extravascular matrix (consistent with fibrin) (Fig. 6A) in lesions of human periodontitis but not in healthy tissues. Furthermore, robust neutrophil infiltrates were observed in the lesions of periodontitis patients (Fig. 6A).

Human genomic data support the role of PLG-dependent pathways in common forms of periodontitis. Recent genome-wide association studies have reported that polymorphisms downstream of *PLG* (rs1247559) are associated with aggressive periodontitis in Europeans (*20*, *21*). In agreement with these previous studies, we found that *PLG* polymorphisms in tandem are associated with severe periodontitis [*PLG* gene–centric $P = 4.1 \times 10^{-4}$; lead single-nucleotide polymorphism (SNP), rs2465836] (Table 1 and Fig. 6B) and with high colonization of the periodontal microbe *Aggregatibacter actinomycetemcomitans* (*Aa*), a pathogen known to be associated with aggressive periodontitis (*Aa* colonization, gene-centric $P = 1.3 \times 10^{-3}$; lead SNP, rs4252200) (Table 1 and Fig. 6C). Mild periodontal disease, gingival inflammation (without periodontal destruction), and colonization with the microbe *Porphyromonas gingivalis* (*Pg*), a pathogen associated with chronic forms of periodontitis, did not yield any material evidence of association with *PLG* polymorphisms (Table 1 and fig. S9). These findings are consistent with previous studies of periodontitis risk–conferring signals in the region immediately downstream of *PLG* (*20*, *21*). Thus, genetic variation in the *PLG* locus is linked with severe and potentially aggressive forms of human periodontitis.

Table 1. PLG common variation is associated with common and potentially aggressive forms of periodontitis. Results of gene-centric association testing of polymorphisms in the *PLG* locus with periodontal clinical and microbiological data. Here, the healthy population serves as a control group. Dashes indicate that no values are being reported in these sections. N/A, not applicable.

Phenotypes	*N*	Gene-centric *P* value*	Lead SNP	[Effect allele] frequency	Lead SNP *P* value
		Clinical periodontal diagnosis			
Healthy	1461	N/A (referent)	–	–	–
Mild disease	812	4.3×10^{-1}	rs14224	[T] 0.60	7.6×10^{-2}
Severe gingival inflammation	160	5.6×10^{-2}	rs783145	[G] 0.47	2.2×10^{-3}
Severe periodontitis†	238	4.1×10^{-4}	rs2465836	[G] 0.83	6.5×10^{-5}
		Subgingival periodontal pathogen colonization			
Aa†	989	1.3×10^{-3}	rs4252200	[A] 0.94	7.5×10^{-5}
Pg	989	4.1×10^{-1}	rs1806449	[T] 0.50	1.1×10^{-1}

*Gene-centric association *P* values obtained using MAGENTA (using a 110 kb upstream to 40 kb downstream window flanking PLG). †Statistically significant gene-centric association after Bonferroni correction for multiple comparisons.

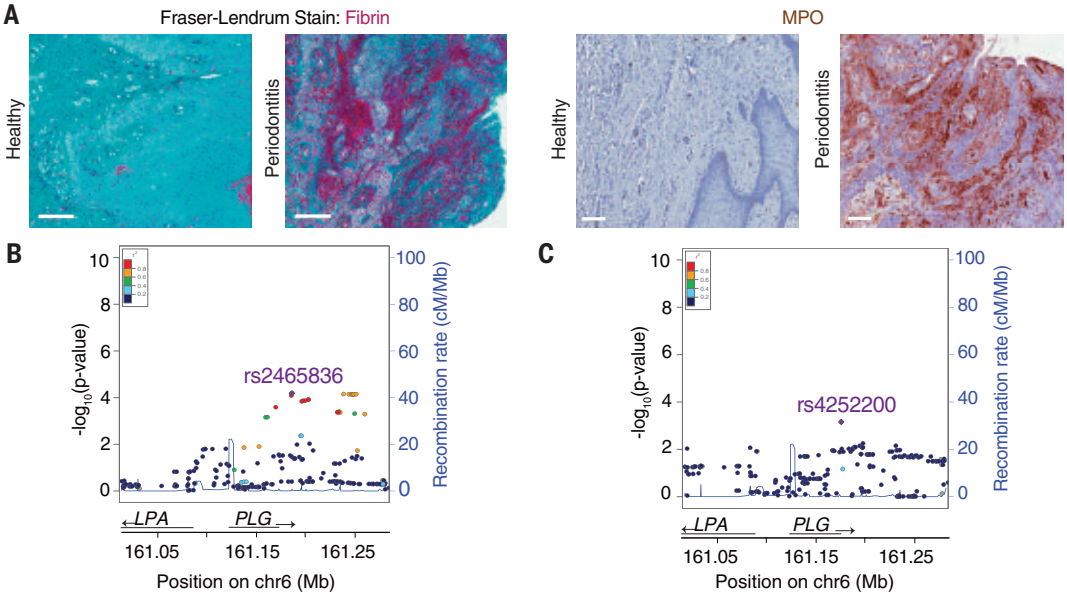


Fig. 6. Fibrin-neutrophil axis as an initiator in common forms of periodontitis. (**A**) Fraser-Lendrum staining for fibrin(ogen) (left) and immunohistochemistry (IHC) staining of MPO (right) in oral mucosal tissues from patients with severe periodontitis and healthy controls. Scale bars, 100 μm. (**B** and **C**) Regional association plots of *PLG* (area ± 110-kb flanking *PLG*) with severe disease (B) and high *Aa* subgingival colonization (C) in the dental atherosclerosis risk in communities study (D-ARIC) population.

Discussion

The homeostatic regulation of neutrophils is critical for optimal function of the immune system within the oral cavity. Both hypo- and hyperrecruitment of neutrophils to the oral mucosa are associated with inflammatory periodontal bone loss. This requirement for finely balanced neutrophil numbers and neutrophil function in the maintenance of periodontal health is revealed by the development of aggressive forms of periodontitis in the setting of both neutropenia and neutrophil accumulation. Congenital disorders associated with defects in neutrophil generation and life cycle, such as the myelokathexis syndrome, severe congenital neutropenia, and leukocyte adhesion deficiency (*18*, *19*, *22*, *23*), lead to early-onset periodontitis, whereas excess neutrophil infiltration and activation has been associated with non-Mendelian forms of periodontal disease (*18*, *19*, *22*, *23*).

The current study provides a clear demonstration that increased neutrophil recruitment per se is insufficient for triggering immunopathology within the oral cavity. Here, we show that the engagement of neutrophils recruited to the periodontal environment with locally deposited fibrin is essential for the full execution of key immunopathology-eliciting neutrophil effector functions, including NETosis. We found that even large increases in the steady-state levels of neutrophils within gingival tissues because of compromised fibrinolysis failed to elicit immunopathology in mice genetically engineered to carry a mutant fibrin unable to engage neutrophils via the major neutrophil integrin, $\alpha_M\beta_2$. Mutations in the $\alpha_M\beta_2$-binding motif protect from fibrin(ogen)-associated immunopathology in models of autoimmune disease, including arthritis and multiple sclerosis, because of the disruption of monocyte, macrophage, and microglial activation (*24*–*26*). However, mechanisms underlying fibrin-triggered mucosal disease and fibrin-neutrophil–mediated immunopathology have not been well understood to date.

Neutrophil recruitment and activation, ROS production, and NETosis have been documented in tissue biopsies of patients with periodontitis (*27*–*32*). Exaggerated activation of neutrophils, ROS production, and release of NETs has been linked to a variety of inflammatory and autoimmune pathologies (*33*–*37*) and has been implicated in inflammatory bone destruction in the setting of arthritis (*38*). However, how neutrophils become activated and participate in periodontal immunopathology is not fully understood. Here, we provide compelling evidence that NETosis is partially responsible for mediating oral mucosal immunopathology and periodontal bone loss in the setting of defective fibrinolysis. Using in vivo models of NET depletion ($Plg^{-/-}$;$Ela^{-/-}$ and $Plg^{-/-}$–DNase I), we confirmed that NETosis plays a significant role in fibrin-induced mucosal immunopathology. However, rescue of bone loss through NET inhibition was not complete, as it was in $Plg^{-/-}$;$Fgg^{390\text{-}396A/390\text{-}396A}$ mice, which suggests additional roles for fibrin-mediated immunopathology beyond NETosis.

We hypothesize that neutrophil activation through the fibrin $\alpha_M\beta_2$ binding site is a physiologically protective mechanism that is in place to activate neutrophils at sites of injury and inflammation and promotes microbial clearance. Yet, in the setting of fibrin accumulation as a result of defective fibrinolysis, we reason that it becomes exaggerated and leads to immunopathology. This mechanism is particularly relevant to the oral mucosa, where neutrophils continuously transmigrate to perform homeostatic functions, even in the setting of health (*32*). However, whether the fibrin-neutrophil axis is responsible for mucosal immunopathology at other barriers (such as the gastrointestinal tract and ocular surface) in *Plg* deficiency is not explored in our study. It is well appreciated that mucosal disease in *Plg* deficiency is mediated by fibrin accumulation in all tissues (*11*), but it is not clear whether $\alpha_M\beta_2$-binding by neutrophils or other location-relevant myeloid subsets or alternative fibrin-mediated inflammatory pathways are engaged to promote immunopathology at different mucosal sites.

We demonstrated that the elimination of the $\alpha_M\beta_2$ binding site on fibrin protects wild-type mice from spontaneous age-related periodontal disease, which indicates that activation of neutrophil effector functions through fibrin engagement could contribute to periodontal disease in individuals with both compromised and normal fibrinolytic function. Human genetic evidence supports the role of *PLG* in common forms of periodontitis, in the absence of a syndromic Mendelian disease. A recent meta-analysis of genome-wide association studies of periodontitis, including almost 15,000 adult participants of European ancestry, has revealed *PLG* polymorphisms as one of the two major risk factors for aggressive periodontitis (*20*). An earlier study has also highlighted polymorphisms downstream of *PLG* as a shared genetic risk locus between cardiovascular disease and periodontitis (*21*). In agreement with this previous work, we have now independently confirmed that *PLG* is a genetic risk factor for periodontitis in a sizeable cohort of adults of European ancestry in North America. Furthermore, we find a strong association of *PLG* polymorphisms with high levels of the periodontal pathogen *A. actinomycetemcomitans* (*Aa*), a microbe typically detected in

patients with aggressive forms of periodontal disease (*39*). The lead SNP in our analysis associated with severe periodontitis (rs2465836) is in strong linkage disequilibrium with the lead marker (rs1247559) in a recent European study of aggressive periodontitis (*20*). Moreover, the SNP (rs4252200) leading the signal for high *Aa* colonization is proximal to the previously reported aggressive periodontitis and cardiovascular disease–related polymorphism (rs4252120) (*40*). Future studies of the functional role of common polymorphisms in the *PLG* locus are necessary to define whether defective fibrinolysis underlies susceptibility and could be targeted in a subset of nonsyndromic periodontitis patients. Notably, the pathway of fibrin-neutrophil–mediated inflammation in periodontal disease uncovered here identifies a putative new therapeutic target in this disease. In support of this notion, a monoclonal antibody that targets the $\alpha_M\beta_2$ binding site on fibrin, γ377 to 395, recently demonstrated therapeutic activity in an animal model of multiple sclerosis (*41*). In conclusion, this work demonstrates an essential role for plasmin-mediated fibrinolysis in oral mucosal homeostasis and disease and suggests that targeting fibrin-neutrophil engagement may be beneficial for the prevention or treatment of the prevalent human disease periodontitis.

Materials and methods

Mice

Plg$^{-/-}$, *Plg*$^{S743A/S743A}$, *Plat*$^{-/-}$;*Plau*$^{-/-}$, *Itgam*$^{-/-}$ ($\alpha_M^{-/-}$), *Fga*$^{-/-}$, *Fgg*$^{390\text{-}396A/390\text{-}396A}$, and *Ela*$^{-/-}$ mice have been previously described (*10*, *11*, *15*, *42*–*46*). All studies used littermate controls. Mice were on a C57BL/6J background, except for *Plg*$^{-/-}$; *Fga*$^{-/-}$ mice and their *Plg*$^{-/-}$;*Fga*$^{+/-}$, *Plg*$^{+/-}$; *Fga*$^{-/-}$, and *Plg*$^{+/-}$;*Fga*$^{+/-}$ littermates, which were on the mixed Black Swiss/C57BL/6J background. Mice were genotyped as previously described (*10*, *11*, *15*, *42*–*45*, *47*). All experiments were performed under approved protocols in an Association for Assessment and Accreditation of Laboratory Animal Care International-certified vivarium and were approved by the Institutional Animal Care and Use Committee before initiation of experimentation. All mice were housed in a SPF facility under regular housing conditions (1 to 5 animals per cage: 14 light–10 dark cycle with standard chow and water ad libitum, temperature of 22°C and 30 to 70% humidity). Mice were provided with water pouches containing ultraviolet (UV)–treated ultra-filtered water, which has been chlorinated to 4 parts per million (ppm). Mice are regularly fed NIH-07 Mouse/Rat Diet (Envigo, cat. no. 7022; Madison, WI). Standard environmental enrichment for all mouse cages was 7099 TEK-Fresh bedding. Singly housed mice also receive a paper Shepherd Shack. Trained animal keepers together with veterinarians ensured the animals' welfare. GF *Plg*$^{-/-}$ mice were generated by Taconic Biosciences, Inc (Derwood, MD).

Study design

Both male and female mice were evaluated together, as our preliminary data did not reveal gender as a confounding factor in periodontitis development. The number of mice used in each experiment is indicated in figures and/or figure legends. The experimental unit was a single mouse. Experiments using genetically modified mice were littermate-controlled. For pharmacological experiments, mice were randomly assigned to control or treatment groups. Mice were tagged in a manner so that the investigator was blinded to genotype and experimental group designation.

Bone loss measurements

The distance between the cementoenamel junction and alveolar bone crest (CEJ-ABC distance) was measured at six predetermined sites after defleshing and staining with methylene blue as previously described (*48*). Results are displayed as means ± SEM.

Micro–computed tomography (micro-CT)

Mouse maxillae were dissected at 24 weeks (six mice for each age group) for quantitative analysis of the alveolar bone. Hemi-maxillae were then examined by a micro-CT system (μCT 50, Scanco Medical AG, Bassersdorf, Switzerland) as previously described (*49*). The values are expressed as means ± SD.

Fraser-Lendrum staining

Decalcification, paraffin embedding, and staining of 5-μm sections of mouse maxillae and mandibles using the Fraser-Lendrum method was performed by Histoserv Inc. (Germantown, MD) (*50*).

Immunofluorescence staining and quantification

Immunofluorescence confocal microscopy was performed using 5-μm sections of decalcified and paraffin-embedded maxillae and mandibles (Histoserv Inc., Germantown, MD). Immunofluorescence staining was performed as previously described (*51*). Primary antibodies were used at 1:1000 dilution [anti-fibrin(ogen): (*42*); anti-Ly6G: BD Pharmingen 551459; anti-MPO: ab65871; anti–Cit-H3: ab5103] O/N at 4°C. Secondary staining was performed with Rhodamine Red-X anti-rat, JacksonImmuno 712-296-150 or Alexa Fluor 647-conjugated AffiniPure F(ab')2 anti-rabbit, JacksonImmuno 711-606-152; 1:500 (Jackson Immuno Research Labs, West Grove, PA) and 4′,6-diamidino-2-phenylindole (DAPI) (1:1000) for 1 hour at room temperature. Inverted Nikon A1R+ confocal microscope (20X-air or 60X-oil immersion) using NIS-Elements software (Nikon, Tokyo, Japan) was used to capture images.

For fibrin quantification, the average fluorescent intensity was measured using NIS-Elements software. A fixed-area square region of interest was drawn around the tissue and the Time Measurement analysis tool was used to measure the average fluorescent intensity within the ROI for the AI555 channel. Five individual fields were measured per tissue and the mean fluorescence intensity values were calculated in Microsoft Excel for Mac (version 16.15). Results are displayed as means ± SEM.

To quantify the MPO- or Cit-H3–stained area in the gingiva, a statistical threshold (set at two standard deviations above the mean of the image histogram) was set. An investigator-drawn region of interest determined the appropriate area of the sample, and the square area was reported. A similar analysis was applied to immunofluorescent images of Ly6G to determine the number of neutrophils in the gingival lesions. The area was calculated and divided by the average area of a neutrophil (determined here to be ~22.1 μm^2, n = 95) to estimate the number of invading neutrophils. The analysis was performed with MetaMorph and results were displayed as means ± SEM.

Antibiotics treatment

Plg$^{-/-}$, *Plat*$^{-/-}$;*Plau*$^{-/-}$, and *Plg*$^{S743A/S743A}$ mice were treated with vancomycin (0.5 mg/ml), doripenem (0.25 mg/ml), and neomycin (5 μg/ml) in autoclaved drinking water from 5-6 to 12-17 weeks of age.

Oral microbiome analysis

Oral murine swab samples were obtained and subjected to DNA isolation, amplicon sequencing of the 16*S* ribosomal RNA (rRNA) V4 region and generated microbiome data were analyzed as previously described (*31*).

RNA-seq

Total RNA was extracted from experimental tissue samples (extracted from M1 to M3 from the mandibles) and sequencing libraries were prepared by the Illumina Nextera XT method following manufacturer's recommendations (Illumina, San Diego, CA). The multiplexed libraries were sequenced in a NextSeq500 instrument in 150-bp paired-end mode. Demultiplexed samples were mapped, and transcripts quantified to the GRCm38.v11 mouse genome using the STAR v2.5.2 aligner. Gene-level counts were filtered to remove low-expressed genes (keeping genes with >5 counts in at least one sample). The filtered gene expression matrix was analyzed by principal component analysis to identify outlier samples (none detected). Differential gene expression was evaluated by three independent statistical methods (DESeq2, edgeR, and Limma-Voom). Differentially expressed genes [false

discover rate (FDR) < 0.01] were considered for further analyses, based on results from DESeq2.

Preparation of single-cell suspensions

Mouse gingivae were dissected and digested for 50 min at 37°C with Collagenase IV (GIBCO, Gaithersburg, MD, USA) and DNase (Sigma, St. Louis, MO, USA), as previously described (*52*).

Flow cytometry

Single-cell suspensions of gingiva were stained with the LIVE/DEAD Fixable Aqua Dead Cell Stain Kit, for 405 nm excitation (L34966; AmCyan; Invitrogen, Waltham, MA), and cell surface markers were stained with the following anti-mouse antibodies at a concentration of 2 μg/ml: anti-Ly6C (AL-21; BV711; BD Biosciences, Franklin Lakes, NJ), anti-Ly6G (1A8; PE-Cy7; BD Biosciences), anti-B220 (RA3-6B2; BV785; BioLegend, San Diego, CA), anti-CD11B (M1/70; PerCP-Cy5.5; BioLegend), anti-TCRβ (H57-597; APC-Cy7; BioLegend), anti-CD11C (N418; FITC; eBioscience), and anti-CD45 (30.F11; BV605; eBioscience). All samples were analyzed using a FACS Fortessa cytometer (BD Biosciences). Data analysis was performed using FlowJo software (Tree Star). Results were displayed as means ± SEM.

Purification of fibrinogen

Wild-type and $Fgg^{390\text{-}396A/390\text{-}396A}$ fibrinogen was purified from citrated-plasma by ammonium sulfate precipitation as described (*53*).

Neutrophil isolation from mouse bone marrow

Neutrophils were isolated as described previously (*54*). Briefly, bone marrow cells were flushed out from femurs and tibiae of 8- to 12-week-old mice with RPMI 1640 supplemented with 2 mM EDTA, 20% heat-inactivated fetal bovine serum (FBS), and 1% penicillin–streptomycin. Neutrophils were isolated from total bone marrow cells by density gradient centrifugation at 300*g* for 30 min without break on Histopaque 1119/Histopaque 1077. The purity of the neutrophils was above 95% as determined by flow cytometry analysis. Then neutrophils were washed with phosphate-buffered saline (PBS) and suspended in RPMI 1640 supplemented with 2% heat-inactivated FBS.

Neutrophil isolation from human blood

Human neutrophils were isolated from heparinized peripheral blood from the NIH blood bank using density gradient centrifugation at 300*g* for 30 min without break on Ficoll-Paque premium (GE Healthcare, Pittsburg, PA) followed by removal of red blood cells (RBCs) using dextran-promoted rosette formation and hypotonic lysis of remaining RBC. For ROS analysis, blood from four female and 11 male donors was used. For NET analyses, blood from four female and three male donors was used.

Cell-based assays

Fibrin coating of wells

Plates were coated with purified wild-type or mutated fibrin (0.60 μg/ml mouse fibrinogen + 5 U/ml mouse thrombin) at 37°C overnight in a humidified incubator. Excess thrombin was washed off with Dulbecco's minimum essential medium (DMEM) and nonspecific binding sites were blocked using 1% nonfat dry milk (1 hour) followed by three DMEM washes. Plates were allowed to dry for 15 min before use.

Cell adhesion assay

Cell adhesion assays were performed in 96-well Black μ-Plates (ibidi USA, Inc., Fitchburg, WI). Aliquots of cells (1×10^5 cells/ml) were added to fibrin-coated wells, incubated at 37°C for 30 min, and subsequently washed with buffered saline to remove nonadherent cells. The number of adherent cells was determined by counting DAPI-stained cells in multiple high-powered (40X) fields. All analyses were performed in triplicate.

ROS production

ROS assays were performed in 96-well clear plates (ThermoFisher Scientific). Aliquots (100 μl) of cells (3.3×10^6 cells/ml) were added to wells and incubated at 37°C for 30 min. Control wells without cells were used to negate any contaminant from fibrin. Cells were supplemented with (25 μl) cytochrome c (1800 nM) with or without 15 nM PMA (phorbol 12-myristate 13-acetate) and/or superoxide dismutase (SOD; 60 U/ml), and absorbance was measured at 550 nm for 5 hour every 10 min. The absorbance values at 550 nm (differences in OD between samples with and without SOD) were converted to nanomoles of O_2^- based on the extinction coefficient of cytochrome c: $D_{E550} = 21 \times 10^3\ M^{-1}cm^{-1}$. The results were expressed as nanomoles of O_2^- per 1×10^6 cells.

NETosis

NETosis assays were performed in eight-well μ-Slides (ibidi USA, Inc.). Aliquots (200 μl) of cells (2.5×10^4 cells/ml) and Hoechst (1 μg/ml) were added per well and incubated at 37°C for 30 min in a humidified incubator. Wells were washed with FluoroBrite DMEM (Thermo Fisher Scientific) to remove nonadherent cells and supplemented with 200 μl of FluoroBrite DMEM containing 0.2 μM Sytox Green. Time-lapse series were acquired every 5 min for 5 hours on an inverted Nikon TI-E microscope using a Plan Fluor 20X (N.A. 0.5) objective and a Flash-4 v3 sCMOS camera (Hamamatsu, Japan). Excitation and emission were provided by a Sola light engine (Lumencor, Beaverton, OR) and standard DAPI and green fluorescent protein (GFP) filter sets (Chroma) for Hoechst and Sytox Green, respectively. Light intensity and camera exposure times varied between 3 and 5% and 100 ms, respectively. An environmental chamber (Precision Plastics) kept samples humidified and at a constant 37°C with 5% CO_2. All hardware was controlled through Nikon Elements. After the first 10 min of imaging, cells were supplemented with 200 μl of FluoroBrite DMEM containing 0.2 μM Sytox Green with or without 15 nM PMA and continued imaging for 5 hours.

NET analysis and quantification

Analysis and quantification were performed as published (*55*). For an example of the processing and analysis, see fig. S7B. Briefly, to distinguish cells undergoing apoptosis from NETosis, a threshold was applied to each image stack and measured using the "Analyze Particles" plugin using different size filters to exclude smaller cells (NETosis > 120; Apoptosis > 52). The Hoechst channel was used to determine live from dead cells. An unsharpen mask (radius = 15, 0.60 scaling) was used to remove out of focus blur. The user then visually determined if each was correct and adjusted false positives before any further statistical analysis.

DNase I treatment

$Plg^{-/-}$ mice of 8 weeks of age were i.p. injected with either 400 U of DNase I (Millipore Sigma, Burlington, MA) in 100 μl of 0.9% NaCl or 0.9% NaCl, twice a week (Tuesday and Friday in the midafternoon) for 12 weeks. Heads were harvested and bone loss measurements were taken as described under "Bone loss measurements."

siRNA-LNP preparation and analysis

siFbg or control siRNA targeting Luciferase (siLuc), purchased from Integrated DNA Technologies (Coralville, IA), were encapsulated in lipid nanoparticles (LNPs) and analyzed as previously described (*56*, *57*). These LNPs naturally accumulate in the liver and are up taken into hepatocytes (*58*), where fibrinogen is made. Briefly, an siRNA solution (amine: phosphate ratio of 3 in 25 mM sodium acetate, pH4) was mixed with a lipid solution consisting of an ionizable lipid (DLin-MC3-DMA), 1,2-distearoyl-sn-glycero-3-phosphocholine (DSPC), cholesterol, and dimyristoyl glycerol-polyethyleneglycol (DMGPEG) (50:10:38.5:1.5 mol%, respectively, in ethanol) using a T-junction mixer and dialyzed against PBS (pH 7.4). Total lipid content and nucleic acid entrapment of the resulting siRNA-LNP was quantified using a Cholesterol E Assay Kit (Wako Chemicals, Richmond, VA) and RiboGreen assay, respectively. We confirmed that siFbg led to a significant decrease in both fibrinogen mRNA in the liver and fibrinogen levels in blood plasma using the mouse total fibrinogen enzyme-linked immunosorbent assay (ELISA) kit (Innovative Research Inc., Arlington, VA).

Fibrin(ogen) depletion in vivo

$Plg^{-/-}$ mice of 8 weeks of age were i.v. injected with either siFbg or siLuc at 2 mg/kg in 100 μl of PBS once every 3 weeks (in the morning) for 12 weeks.

Human oral mucosal samples

Gingival biopsies were harvested from healthy volunteers and patients with severe untreated periodontal disease. All patients were enrolled in an IRB-approved clinical study at the NIH hospital (NCT no. 01568697) and provided informed written consent. Human biopsies were stained for MPO as follows and fibrin (Fraser-Lendrum stain) as described above. Antigen retrieval in 10 mM citric acid buffer was used, and sections were blocked with 5% bovine serum albumin in PBS for 1 hour. Sections were incubated with MPO antibodies (ab9535) at 1:500 dilution O/N, followed by secondary anti-rabbit antibody (1:500; ImmPRESS HRP Reagent kit, Vector Laboratories) for 30 min. Sections were then washed three times in PBS followed by the addition of developing solution (ImmPACT DAB, Vector Laboratories). The reaction was stopped with water and slides counterstained with hematoxylin QS (Vector Laboratories Inc., Burlingame, CA).

Human genetic associations with clinical and microbiological periodontal disease parameters

To examine the association of polymorphisms in the *PLG* gene locus with the presence and severity of human periodontal disease in an independent cohort, we used existing clinical, microbiological, and genotype information from the Atherosclerosis Risk in Communities (ARIC) study (*59*). ARIC is a prospective cohort study of atherosclerosis and cardiovascular disease risk factors which included a Dental ARIC ancillary study with comprehensive oral examinations in a subset of ARIC participants [n = 4504 European Americans (EA); mean age = 62 years, females = 52%]. The analysis of existing, deidentified genotype and clinical periodontal data was approved as nonhuman subjects research (NHSR) by the UNC-Chapel Hill IRB (no. 11-1035). All participants of the parent ARIC study and ancillary studies provided written informed consent. Oral microbiome samples were collected from subgingival tooth surfaces in a subset of Dental ARIC participants (n = 1020 EA) and were subsequently analyzed for the presence and abundance of eight common periodontal pathogens. In this study, we based oral disease diagnoses on the recently introduced P^3 classification system (*60*), separating the population into seven distinct oral health categories. Four categories focused on the severity of periodontal disease [mild disease, severe inflammation (inflammation without periodontal bone loss), localized disease, and severe periodontitis (severe periodontal bone loss)]. Two categories reflected oral disease burden not specific to periodontitis (tooth loss and severe tooth loss), and the final category was designated for periodontal health (*61*, *62*). Microbiome-specific categories were assigned for patients with high [i.e., in the top quintile (*63*)] detection of either *Porphyromonas gingivalis* (*Pg*, associated with the most common, chronic types of periodontitis) (*64*) or *Aggregatibacter actinomycetemcomitans* (*Aa*, associated with aggressive periodontitis) (*39*).

We used existing genotype data that were initially generated using the Affymetrix Genome-Wide Human SNP Array 6.0 chip (n = ~900,000 markers for SNPs) and then imputed to the HapMap Phase II CEU build 36 panel resulting in a final analysis set of 2.1 million SNPs with minor allele frequency of >5% (*59*). Genetic association models for single markers (i.e., SNPs) were based on logistic regression (i.e., "disease" versus "healthy" periodontal groups and "high" versus "lower" pathogen colonization), assuming log-additive allelic effects, and including adjustment terms of study design, 10 principal components for ancestry, age, and sex.

To evaluate *PLG* genetic association with clinical and microbiological parameters of periodontal disease, first we defined an extended gene region around it, 110 kb upstream to 40 kb downstream of the gene boundaries, harboring 153 SNPs. The selection of the specific boundaries was based upon simulation and actual data applications of MAGENTA (*65*) that were used to obtain gene-centric association estimates. This was done by synthesizing single-marker association signals in the defined *PLG* gene region into a single association estimate for *PLG* (Z score and corresponding gene-centric P value), accounting for the number of SNPs tested and their local linkage disequilibrium (LD) structure, based on the human genome build 36. Here, we obtained *PLG*-centric estimates of association with clinical diagnoses and microbiological periodontal parameters and report gene P values, as well as information for the lead (i.e., lowest P value) SNP in the gene region. In this study, we formally tested for five *PLG*-trait associations (mild periodontitis, severe gingival inflammation, severe periodontitis, high *Aa*, and high *Pg* colonization). Because the P^3 system includes six clinical disease groups (including localized disease and patterns of tooth loss), and because four periodontal pathogen colonization traits have been previously reported (*63*), we applied a conservative Bonferroni multiple testing correction assuming 10 independent tests (critical value $P < 5 \times 10^{-3}$). Finally, to illustrate and provide more information about the genomic context of the identified *PLG* locus associations, we generated regional association (Locus Zoom) plots (*66*) using single-marker association results in areas ±110 kb flanking *PLG*.

Statistical analysis

For in vivo assays, unpaired Student's t test, or one-way analysis of variance (ANOVA) (when >2 groups) was performed. For in vitro assays with human neutrophils, results were displayed as means ± SEM, and two-tailed Wilcoxon signed-rank test was performed. For in vitro assays with mouse neutrophils, results were displayed as means ± SEM, and Wilcoxon signed-rank test was performed for paired values and Mann-Whitney U test for unpaired samples. For time point analyses (ROS and NETosis), linear regression analysis was performed to test whether differences in slopes were significantly different. Confidence interval lines (95%) are displayed on graphs. All statistical analyses were performed on Graph Pad Prism (v7). P values for principal coordinates analysis (PCoA) plot of global microbial community structure of oral samples from *Plg* mice were determined using analysis of molecular variance (AMOVA). Microbial diversity was measured through the nonparametric version of the Shannon diversity index of oral samples. P values were determined using Mann-Whitney U test. Statistical significance of microbiome composition at the operational taxonomic unit (OTU) level was determined via LEfSe (*67*).

REFERENCES AND NOTES

1. K. Tefs *et al*., Molecular and clinical spectrum of type I plasminogen deficiency: A series of 50 patients. *Blood* **108**, 3021–3026 (2006). doi: 10.1182/blood-2006-04-017350; pmid: 16849641
2. V. Schuster, B. Hügle, K. Tefs, Plasminogen deficiency. *J. Thromb. Haemost.* **5**, 2315–2322 (2007). doi: 10.1111/j.1538-7836.2007.02776.x; pmid: 17900274
3. J. Frimodt-Møller, Conjunctivitis ligneosa combined with a dental affection. Report of a case. *Acta Ophthalmol.* **51**, 34–38 (1973). doi: 10.1111/j.1755-3768.1973.tb08243.x; pmid: 4739674
4. I. Kurtulus Waschulewski *et al*., Immunohistochemical analysis of the gingiva with periodontitis of type I plasminogen deficiency compared to gingiva with gingivitis and periodontitis and healthy gingiva. *Arch. Oral Biol.* **72**, 75–86 (2016). doi: 10.1016/j.archoralbio.2016.07.013; pmid: 27552374
5. U. Ertas, N. Saruhan, O. Gunhan, Ligneous periodontitis in a child with plasminogen deficiency. *Niger. J. Clin. Pract.* **20**, 1656–1658 (2017). doi: 10.4103/1119-3077.224133; pmid: 29379003
6. S. H. Neering *et al*., Periodontitis associated with plasminogen deficiency: A case report. *BMC Oral Health* **15**, 59 (2015). doi: 10.1186/s12903-015-0045-3; pmid: 25971786
7. M. El-Darouti, A. A. Zayed, G. Y. El-Kamah, M. I. Mostafa, Ligneous conjunctivitis with oral mucous membrane involvement and decreased plasminogen level. *Pediatr. Dermatol.* **26**, 448–451 (2009). doi: 10.1111/j.1525-1470.2009.00951.x; pmid: 19689523
8. R. P. Darveau, Periodontitis: A polymicrobial disruption of host homeostasis. *Nat. Rev. Microbiol.* **8**, 481–490 (2010). doi: 10.1038/nrmicro2337; pmid: 20514045
9. G. Hajishengallis, Periodontitis: From microbial immune subversion to systemic inflammation. *Nat. Rev. Immunol.* **15**, 30–44 (2015). doi: 10.1038/nri3785; pmid: 25534621
10. T. H. Bugge, M. J. Flick, C. C. Daugherty, J. L. Degen, Plasminogen deficiency causes severe thrombosis but is compatible with development and reproduction. *Genes Dev.* **9**, 794–807 (1995). doi: 10.1101/gad.9.7.794; pmid: 7705657
11. T. H. Bugge *et al*., Loss of fibrinogen rescues mice from the pleiotropic effects of plasminogen deficiency. *Cell* **87**, 709–719 (1996). doi: 10.1016/S0092-8674(00)81390-2; pmid: 8929539
12. R. Sulniute, T. Lindh, M. Wilczynska, J. Li, T. Ny, Plasmin is essential in preventing periodontitis in mice. *Am. J. Pathol.* **179**, 819–828 (2011). doi: 10.1016/j.ajpath.2011.05.003; pmid: 21704601

13. A. F. Drew *et al*., Ligneous conjunctivitis in plasminogen-deficient mice. *Blood* **91**, 1616–1624 (1998). doi: 10.1182/blood.V91.5.1616; pmid: 9473227
14. S. Liang, K. B. Hosur, H. Domon, G. Hajishengallis, Periodontal inflammation and bone loss in aged mice. *J. Periodontal Res.* **45**, 574–578 (2010). doi: 10.1111/j.1600-0765.2009.01245.x; pmid: 20337897
15. M. J. Flick *et al*., Leukocyte engagement of fibrin(ogen) via the integrin receptor $\alpha_M\beta_2$/Mac-1 is critical for host inflammatory response in vivo. *J. Clin. Invest.* **113**, 1596–1606 (2004). doi: 10.1172/JCI20741; pmid: 15173886
16. V. Papayannopoulos, K. D. Metzler, A. Hakkim, A. Zychlinsky, Neutrophil elastase and myeloperoxidase regulate the formation of neutrophil extracellular traps. *J. Cell Biol.* **191**, 677–691 (2010). doi: 10.1083/jcb.201006052; pmid: 20974816
17. A. Akinc *et al*., The Onpattro story and the clinical translation of nanomedicines containing nucleic acid-based drugs. *Nat. Nanotechnol.* **14**, 1084–1087 (2019). doi: 10.1038/s41565-019-0591-y; pmid: 31802031
18. E. Hajishengallis, G. Hajishengallis, Neutrophil homeostasis and periodontal health in children and adults. *J. Dent. Res.* **93**, 231–237 (2014). doi: 10.1177/0022034513507956; pmid: 24097856
19. G. Hajishengallis, T. Chavakis, E. Hajishengallis, J. D. Lambris, Neutrophil homeostasis and inflammation: Novel paradigms from studying periodontitis. *J. Leukoc. Biol.* **98**, 539–548 (2015). doi: 10.1189/jlb.3VMR1014-468R; pmid: 25548253
20. M. Munz *et al*., Meta-analysis of genome-wide association studies of aggressive and chronic periodontitis identifies two novel risk loci. *Eur. J. Hum. Genet.* **27**, 102–113 (2019). doi: 10.1038/s41431-018-0265-5; pmid: 30218097
21. A. S. Schaefer *et al*., Genetic evidence for *PLASMINOGEN* as a shared genetic risk factor of coronary artery disease and periodontitis. *Circ. Cardiovasc. Genet.* **8**, 159–167 (2015). doi: 10.1161/CIRCGENETICS.114.000554; pmid: 25466412
22. N. M. Moutsopoulos *et al*., Defective neutrophil recruitment in leukocyte adhesion deficiency type I disease causes local IL-17–driven inflammatory bone loss. *Sci. Transl. Med.* **6**, 229ra40 (2014). doi: 10.1126/scitranslmed.3007696; pmid: 24670684
23. L. M. Silva, L. Brenchley, N. M. Moutsopoulos, Primary immunodeficiencies reveal the essential role of tissue neutrophils in periodontitis. *Immunol. Rev.* **287**, 226–235 (2019). doi: 10.1111/imr.12724; pmid: 30565245
24. M. J. Flick *et al*., Fibrin(ogen) exacerbates inflammatory joint disease through a mechanism linked to the integrin $\alpha_M\beta_2$ binding motif. *J. Clin. Invest.* **117**, 3224–3235 (2007). doi: 10.1172/JCI30134; pmid: 17932565
25. R. A. Adams *et al*., The fibrin-derived $\gamma^{377\text{-}395}$ peptide inhibits microglia activation and suppresses relapsing paralysis in central nervous system autoimmune disease. *J. Exp. Med.* **204**, 571–582 (2007). doi: 10.1084/jem.20061931; pmid: 17339406
26. D. Davalos *et al*., Fibrinogen-induced perivascular microglial clustering is required for the development of axonal damage in neuroinflammation. *Nat. Commun.* **3**, 1227 (2012). doi: 10.1038/ncomms2230; pmid: 23187627
27. J. B. Matthews, H. J. Wright, A. Roberts, P. R. Cooper, I. L. Chapple, Hyperactivity and reactivity of peripheral blood neutrophils in chronic periodontitis. *Clin. Exp. Immunol.* **147**, 255–264 (2007). doi: 10.1111/j.1365-2249.2006.03276.x; pmid: 17223966
28. J. B. Matthews *et al*., Neutrophil hyper-responsiveness in periodontitis. *J. Dent. Res.* **86**, 718–722 (2007). doi: 10.1177/154405910708600806; pmid: 17652198
29. E. A. Nicu, P. Rijkschroeff, E. Wartewig, K. Nazmi, B. G. Loos, Characterization of oral polymorphonuclear neutrophils in periodontitis patients: A case-control study. *BMC Oral Health* **18**, 149 (2018). doi: 10.1186/s12903-018-0615-2; pmid: 30143044
30. P. C. White, I. J. Chicca, P. R. Cooper, M. R. Milward, I. L. Chapple, Neutrophil Extracellular Traps in Periodontitis: A Web of Intrigue. *J. Dent. Res.* **95**, 26–34 (2016). doi: 10.1177/0022034515609097; pmid: 26442948
31. N. Dutzan *et al*., A dysbiotic microbiome triggers T_H17 cells to mediate oral mucosal immunopathology in mice and humans. *Sci. Transl. Med.* **10**, eaat0797 (2018). doi: 10.1126/scitranslmed.aat0797; pmid: 30333238
32. D. W. Williams *et al*., Human oral mucosa cell atlas reveals a stromal-neutrophil axis regulating tissue immunity. *Cell* **184**, 4090–4104.e15 (2021). doi: 10.1016/j.cell.2021.05.013; pmid: 34129837
33. B. Amulic, C. Cazalet, G. L. Hayes, K. D. Metzler, A. Zychlinsky, Neutrophil function: From mechanisms to disease. *Annu. Rev. Immunol.* **30**, 459–489 (2012). doi: 10.1146/annurev-immunol-020711-074942; pmid: 22224774
34. E. Kolaczkowska, P. Kubes, Neutrophil recruitment and function in health and inflammation. *Nat. Rev. Immunol.* **13**, 159–175 (2013). doi: 10.1038/nri3399; pmid: 23435331
35. K. Ley *et al*., Neutrophils: New insights and open questions. *Sci. Immunol.* **3**, eaat4579 (2018). doi: 10.1126/sciimmunol.aat4579; pmid: 30530726
36. G. Sollberger, D. O. Tilley, A. Zychlinsky, Neutrophil Extracellular Traps: The Biology of Chromatin Externalization. *Dev. Cell* **44**, 542–553 (2018). doi: 10.1016/j.devcel.2018.01.019; pmid: 29533770
37. F. V. S. Castanheira, P. Kubes, Neutrophils and NETs in modulating acute and chronic inflammation. *Blood* **133**, 2178–2185 (2019). doi: 10.1182/blood-2018-11-844530; pmid: 30898862
38. C. Carmona-Rivera *et al*., Synovial fibroblast-neutrophil interactions promote pathogenic adaptive immunity in rheumatoid arthritis. *Sci. Immunol.* **2**, eaag3358 (2017). doi: 10.1126/sciimmunol.aag3358; pmid: 28649674
39. B. Schacher *et al*., *Aggregatibacter actinomycetemcomitans* as indicator for aggressive periodontitis by two analysing strategies. *J. Clin. Periodontol.* **34**, 566–573 (2007). doi: 10.1111/j.1600-051X.2007.01080.x; pmid: 17433043
40. N. L. Overton, A. Simpson, P. Bowyer, D. W. Denning, Genetic susceptibility to severe asthma with fungal sensitization. *Int. J. Immunogenet.* **44**, 93–106 (2017). doi: 10.1111/iji.12312; pmid: 28371335
41. J. K. Ryu *et al*., Fibrin-targeting immunotherapy protects against neuroinflammation and neurodegeneration. *Nat. Immunol.* **19**, 1212–1223 (2018). doi: 10.1038/s41590-018-0232-x; pmid: 30323343
42. T. T. Suh *et al*., Resolution of spontaneous bleeding events but failure of pregnancy in fibrinogen-deficient mice. *Genes Dev.* **9**, 2020–2033 (1995). doi: 10.1101/gad.9.16.2020; pmid: 7649481
43. P. Carmeliet *et al*., Physiological consequences of loss of plasminogen activator gene function in mice. *Nature* **368**, 419–424 (1994). doi: 10.1038/368419a0; pmid: 8133887
44. A. Coxon *et al*., A novel role for the β2 integrin CD11b/CD18 in neutrophil apoptosis: A homeostatic mechanism in inflammation. *Immunity* **5**, 653–666 (1996). doi: 10.1016/S1074-7613(00)80278-2; pmid: 8986723
45. T. Iwaki *et al*., The generation and characterization of mice expressing a plasmin-inactivating active site mutation. *J. Thromb. Haemost.* **8**, 2341–2344 (2010). doi: 10.1111/j.1538-7836.2010.03995.x; pmid: 20653841
46. A. Belaaouaj *et al*., Mice lacking neutrophil elastase reveal impaired host defense against gram negative bacterial sepsis. *Nat. Med.* **4**, 615–618 (1998). doi: 10.1038/nm0598-615; pmid: 9585238
47. T. H. Bugge *et al*., Urokinase-type plasminogen activator is effective in fibrin clearance in the absence of its receptor or tissue-type plasminogen activator. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 5899–5904 (1996). doi: 10.1073/pnas.93.12.5899; pmid: 8650190
48. M. A. Eskan *et al*., The leukocyte integrin antagonist Del-1 inhibits IL-17-mediated inflammatory bone loss. *Nat. Immunol.* **13**, 465–473 (2012). doi: 10.1038/ni.2260; pmid: 22447028
49. A. Wilensky, Y. Gabet, H. Yumoto, Y. Houri-Haddad, L. Shapira, Three-dimensional quantification of alveolar bone loss in *Porphyromonas gingivalis*-infected mice using micro-computed tomography. *J. Periodontol.* **76**, 1282–1286 (2005). doi: 10.1902/jop.2005.76.8.1282; pmid: 16101359
50. A. C. Lendrum, D. S. Fraser, W. Slidders, R. Henderson, Studies on the character and staining of fibrin. *J. Clin. Pathol.* **15**, 401–413 (1962). doi: 10.1136/jcp.15.5.401; pmid: 13929601
51. L. M. Silva *et al*., Plasmin-mediated fibrinolysis enables macrophage migration in a murine model of inflammation. *Blood* **134**, 291–303 (2019). doi: 10.1182/blood.2018874859; pmid: 31101623
52. N. Dutzan, L. Abusleme, J. E. Konkel, N. M. Moutsopoulos, Isolation, Characterization and Functional Examination of the Gingival Immune Cell Network. *J. Vis. Exp.* **108**, e53736 (2016). doi: 10.3791/53736; pmid: 26967370
53. J. M. Prasad *et al*., Mice expressing a mutant form of fibrinogen that cannot support fibrin formation exhibit compromised antimicrobial host defense. *Blood* **126**, 2047–2058 (2015). doi: 10.1182/blood-2015-04-639849; pmid: 26228483
54. M. Swamydas, M. S. Lionakis, Isolation, purification and labeling of mouse bone marrow neutrophils for functional studies and adoptive transfer experiments. *J. Vis. Exp.* **77**, e50586 (2013). doi: 10.3791/50586; pmid: 23892876
55. L. M. Silva, N. Moutsopoulos, T. H. Bugge, A. Doyle, Live Imaging and Quantification of Neutrophil Extracellular Trap Formation. *Curr Protoc* **1**, e157 (2021). doi: 10.1002/cpz1.157; pmid: 34260822
56. A. W. Strilchuk *et al*., Sustained depletion of FXIII-A by inducing acquired FXIII-B deficiency. *Blood* **136**, 2946–2954 (2020). doi: 10.1182/blood.2020004976; pmid: 32678423
57. J. A. Kulkarni *et al*., On the Formation and Morphology of Lipid Nanoparticles Containing Ionizable Cationic Lipids and siRNA. *ACS Nano* **12**, 4787–4795 (2018). doi: 10.1021/acsnano.8b01516; pmid: 29614232
58. A. Akinc *et al*., Targeted delivery of RNAi therapeutics with endogenous and exogenous ligand-based mechanisms. *Mol. Ther.* **18**, 1357–1364 (2010). doi: 10.1038/mt.2010.85; pmid: 20461061
59. K. Divaris *et al*., Exploring the genetic basis of chronic periodontitis: A genome-wide association study. *Hum. Mol. Genet.* **22**, 2312–2324 (2013). doi: 10.1093/hmg/ddt065; pmid: 23459936
60. J. D. Beck, K. L. Moss, T. Morelli, S. Offenbacher, In search of appropriate measures of periodontal status: The Periodontal Profile Phenotype (P^3) system. *J. Periodontol.* **89**, 166–175 (2018). doi: 10.1002/JPER.17-0424; pmid: 29520827
61. T. Morelli *et al*., Derivation and Validation of the Periodontal and Tooth Profile Classification System for Patient Stratification. *J. Periodontol.* **88**, 153–165 (2017). doi: 10.1902/jop.2016.160379; pmid: 27620653
62. K. Divaris, K. Moss, J. D. Beck, Biologically informed stratification of periodontal disease holds the key to achieving precision oral health. *J. Periodontol.* **91**, S50–S55 (2020). doi: 10.1002/JPER.20-0096; pmid: 32432812
63. K. Divaris *et al*., Genome-wide association study of periodontal pathogen colonization. *J. Dent. Res.* **91**, S21–S28 (2012). doi: 10.1177/0022034512447951; pmid: 22699663
64. R. P. Darveau, G. Hajishengallis, M. A. Curtis, *Porphyromonas gingivalis* as a potential community activist for disease. *J. Dent. Res.* **91**, 816–820 (2012). doi: 10.1177/0022034512453589; pmid: 22772362
65. A. V. Segrè, DIAGRAM Consortium, MAGIC investigators, L. Groop, V. K. Mootha, M. J. Daly, D. Altshuler, Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. *PLOS Genet.* **6**, e1001058 (2010). doi: 10.1371/journal.pgen.1001058; pmid: 20714348
66. R. J. Pruim *et al*., LocusZoom: Regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336–2337 (2010). doi: 10.1093/bioinformatics/btq419; pmid: 20634204
67. N. Segata *et al*., Metagenomic biomarker discovery and explanation. *Genome Biol.* **12**, R60 (2011). doi: 10.1186/gb-2011-12-6-r60; pmid: 21702898

ACKNOWLEDGMENTS

We thank M. J. Danton for critically reviewing this manuscript. We also thank the NCI microbiome sequencing core for assistance with 16S rRNA sequencing and data deposition. **Funding:** This work was funded by the NIH NIDCR Intramural Research Program (T.H.B. and N.M.M.); NIDCR 1K99DE030124-01A1 (L.M.S.); NIDCR Veterinary Resources Core (ZIC DE000740-05); NIDCR/NIDCD Genomics and Computational Biology Core (ZIC DC000086); NIDCR Combined Technical Research Core (ZIC DE000729-09); and NIDCR U01DE025046 (C.S.A. and K.D.). This work was also supported in part by NIH R01DK112778 and NIH R01CA211098 (M.J.F.), NIH R01HL013423-43A1 (F.J.C.), and FONDECYT grant 11180505 (L.A.). **Author contributions:** Conceptualization: L.M.S., T.H.B., and N.M.M.; Formal analysis: L.M.S., D.M., K.M., C.S.A., and K.D.; Funding acquisition: T.H.B. and N.M.M.; Investigation: L.M.S., A.D.D., T.G.-W., C.L.T., L.A., and C.E.Z.; Methodology: L.M.S., A.D.D., T.G.-W., N.D., C.L.T., L.A., E.M.C., A.G.L., C.E.Z., M.S., M.S.L., and M.J.F.; Project administration: L.M.S., T.H.B., and N.M.M.; Resources: L.J.J., J.L., P.J., V.K., F.J.C., C.J.K., and M.J.F.; Software: A.D.D.; Visualization: L.M.S.; Writing – original draft: L.M.S., A.D.D., K.D., T.H.B., and N.M.M.; Writing – review and editing: all authors. **Competing interests:** C.J.K. is a director and shareholder of NanoVation Therapeutics Inc., which is commercializing RNA-based therapies. C.J.K., L.J.J., and J.L. are inventors on pending intellectual property related to RNA-based therapies. The other authors declare that they have no competing interests. **Data and materials availability:** RNA-seq data have been deposited to the GEO database (record GSE178616) and can be accessed using the token: ylkhyomevlszjuv. Sequencing data have been deposited to the Sequence Read Archive (SRA) with the BioProjectID PRJNA753711 (https://dataview.ncbi.nlm.nih.gov/object/PRJNA753711?reviewer=a5barpkgib249d5v2res6mtgef). All other data are available in the main text or supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abl5450
Figs. S1 to S9
Tables S1 and S2
MDAR Reproducibility Checklist

21 July 2021; accepted 29 October 2021
10.1126/science.abl5450

Science
JOURNALS AAAS



Publish your research in the *Science* family of journals

The *Science* family of journals (*Science, Science Advances, Science Immunology, Science Robotics, Science Signaling*, and *Science Translational Medicine)* are among the most highly-regarded journals in the world for quality and selectivity. Our peer-reviewed journals are committed to publishing cutting-edge research, incisive scientific commentary, and insights on what's important to the scientific world at the highest standards.

Submit your research today!

Learn more at **Science.org/journals**

RESEARCH ARTICLE SUMMARY

HUMAN EVOLUTION

The energetics of uniquely human subsistence strategies

Thomas S. Kraft*, Vivek V. Venkataraman, Ian J. Wallace, Alyssa N. Crittenden, Nicholas B. Holowka, Jonathan Stieglitz, Jacob Harris, David A. Raichlen, Brian Wood, Michael Gurven†, Herman Pontzer*†

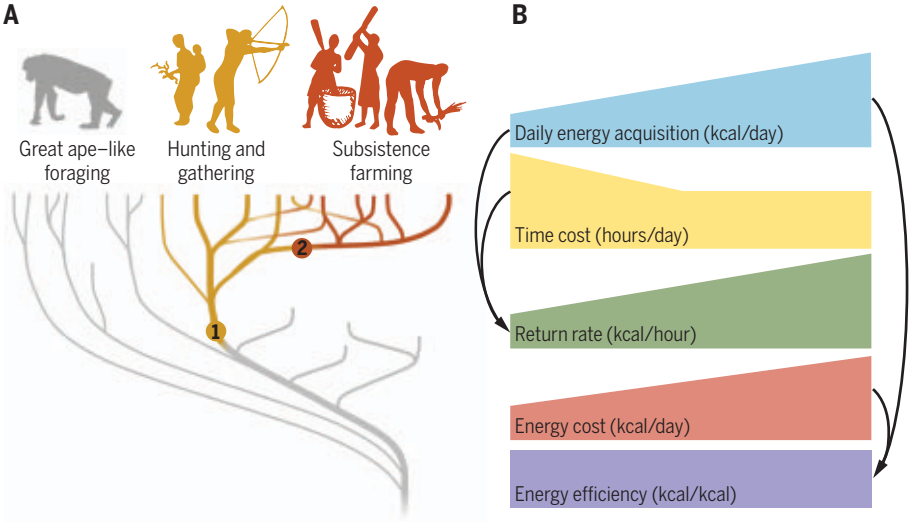
INTRODUCTION: Relative to other great apes, humans have large brains, long life spans, higher fertility and larger neonates, and protracted periods of childhood dependency and development. Although these traits constitute the unique human life history that underlies the ecological success of our species, they also require human adults to meet extraordinarily high energetic demands. Determining how human subsistence strategies have met such extreme energy needs, given time and energy expenditure constraints, is thus key to understanding the origins of derived human traits.

RATIONALE: Two major transitions in hominin subsistence strategies are thought to have elevated energy capture: (i) the development of hunting and gathering ~2.5 million years ago, which coincided with brain enlargement and extended postnatal growth, and (ii) the rise of agriculture ~12,000 years ago, which was accompanied by substantial increases in fertility and population densities. These transitions are associated with the exploitation of novel food sources, but it is not clear how the energy and time budgets of early human foragers and farmers shifted to accommodate expensive traits. Some evolutionary reconstructions contend that economical locomotion, cooperation, the use of sophisticated tools, and eventually agriculture increased energy efficiency (i.e., energy gained versus energy spent), beyond that of other great apes. Alternatively, unique human subsistence strategies may reduce time and improve yield, increasing return rates (i.e., energy gained versus time spent).

To test these ideas, we compared subsistence costs (energy and time) and energy acquisition among wild orangutans, gorillas, and chimpanzees with high-resolution data on total energy expenditure, food acquisition, and time allocation, collected among Tanzanian hunter-gatherers (Hadza) and Bolivian forager-horticulturalists (Tsimane). Both populations actively forage (hunt, gather), whereas the Tsimane also practice slash-and-burn horticulture, which permits exploration of further changes in the energetics of subsistence associated with farming. We also assembled a global subsistence energetics database of contemporary hunter-gatherers and horticulturalists.

RESULTS: Relative to other great apes, human hunter-gatherers and horticulturalists spend more energy daily on subsistence, and they achieve similar energy efficiencies despite having more economical locomotion and using sophisticated technologies. In contrast, humans attain much greater return rates, spending less time on subsistence while acquiring more energy per hour. Further, horticulture is associated with higher return rates than hunting and gathering, despite minimal differences in the amount of time devoted to subsistence. Findings from our detailed study of the Hadza and Tsimane were consistent with those from the larger cross-cultural database of subsistence-level societies. Together, these results support prior evidence that the adoption of farming could have been motivated by greater gains per time spent working, and refute the notion that farming lifestyles are necessarily associated with increased labor time.

CONCLUSION: These findings revise our understanding of human energetics and evolution, indicating that humans afford expanded energy budgets primarily by increasing rates of energy acquisition, and not through energy-saving adaptations (such as economical bipedalism or sophisticated tool use) that decrease overall costs. Relative to other great apes, human subsistence strategies are characterized by high-intensity, high-cost extractive activities and expanded day ranges that provide more calories in less time. These results suggest that energy gained from improvements in efficiency throughout human evolution were primarily channeled toward further increasing foraging intensity rather than reducing the energetic costs of subsistence. Greater energetic gains per unit time are the reward for humans' intense and behaviorally sophisticated subsistence strategies. Humans' high-cost but high-return strategy is ecologically risky, and we argue that it was only possible in the context of increased cooperation, intergenerational food sharing, and a division of labor. We contend that the time saved by human subsistence strategies provided more leisure time for social interaction and social learning in central-place locations, which is critical for cumulative cultural evolution. ■



Major transitions in hominoid subsistence energetics. (**A**) The shift from great ape–like foraging to hunting and gathering (1) and the adoption of subsistence farming during the Neolithic Revolution (2) involved changes in behavior and technology to allow access to novel food resources. (**B**) Through these transitions, humans paid higher energy costs in order to acquire a greater number of calories in less time; transitions from left to right are as depicted in (A). Human subsistence minimizes time costs but not energy costs, resulting in improved return rates but efficiency similar to that of other great apes.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: tkraft@anth.ucsb.edu (T.S.K.); herman.pontzer@duke.edu (H.P.)
†These authors contributed equally to this work.
Cite this article as T. S. Kraft *et al.*, *Science* **374**, eabf0130 (2021). DOI: 10.1126/science.abf0130

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abf0130

ILLUSTRATIONS: SAMANTHA SHIELDS

RESEARCH ARTICLE

HUMAN EVOLUTION

The energetics of uniquely human subsistence strategies

Thomas S. Kraft[1,2,3]*, Vivek V. Venkataraman[4,5], Ian J. Wallace[6], Alyssa N. Crittenden[7], Nicholas B. Holowka[8], Jonathan Stieglitz[4], Jacob Harris[9,10], David A. Raichlen[11], Brian Wood[2,9], Michael Gurven[1]†, Herman Pontzer[12,13]*†

The suite of derived human traits, including enlarged brains, elevated fertility rates, and long developmental periods and life spans, imposes extraordinarily high energetic costs relative to other great apes. How do human subsistence strategies accommodate our expanded energy budgets? We found that relative to other great apes, human hunter-gatherers and subsistence farmers spend more energy but less time on subsistence, acquire substantially more energy per hour, and achieve similar energy efficiencies. These findings revise our understanding of human energetic evolution by indicating that humans afford expanded energy budgets primarily by increasing rates of energy acquisition, not through energy-saving adaptations such as economical bipedalism or sophisticated tool use that decrease subsistence costs and improve the energetic efficiency of subsistence. We argue that the time saved by human subsistence strategies provides more leisure time for social interaction and social learning in central-place locations and would have been critical for cumulative cultural evolution.

In an evolutionary context, the energetic demands faced by adult humans are extraordinarily high. Relative to other hominoids, humans exhibit metabolic acceleration (i.e., elevated body size–adjusted energy expenditure) associated with the high energetic costs of maintaining large brains, a long lifespan, and high reproductive investment (both mass and number of offspring per year) (*1*). Additionally, adult humans must produce a caloric surplus to subsidize a long period of childhood dependency of their offspring (*2*). A major transition in hominin subsistence behavior was likely critical for accommodating increased energetic demands that accompanied changes in morphology, brain size, and life history ~1.5 million to 2.5 million years ago, the time period during which central-place hunting and gathering likely arose (*3*). Likewise, the rise of agriculture represents a marked shift in human subsistence that coincided with substantial increases in human reproductive rates and population densities (*4*).

Although these subsistence strategies have elevated the scale of daily human energy acquisition beyond that of other great apes (*2*), it remains poorly understood how humans manage the potentially severe energy and time costs of obtaining food through hunting and gathering or subsistence agriculture (*5*). Energy expenditure is limited by daily food intake and the amount of energy stored in the body (primarily as fat), whereas time is limited by the number of hours in a day. Both energy and time spent on subsistence activities entail opportunity costs, including energy that could otherwise be expended on bodily maintenance, reproductive investment, or provisioning kin, and time that could be invested in socialization or mating effort. Energy and time investments in subsistence therefore depend on the extent to which these currencies are limiting and the relative fitness benefits of investing those resources elsewhere (*5*–*7*).

Organisms typically pursue two alternative strategies to increase energy availability. One strategy involves increasing energy efficiency. Although efficiency has been inconsistently defined in reference to time or energy costs in the anthropological literature [(*8*), p. 185], here we follow convention from foraging ecology in referring to efficiency as the ratio of energy gained relative to energy spent on subsistence (*5*, *9*, *10*). A second strategy is to increase the net acquisition rate, the amount of energy acquired minus energy spent divided by the time spent on subsistence (*10*). Theoretical work demonstrates that the conditions under which energy efficiency or net acquisition rate are maximized depend on the interplay between the type of foraging (e.g., feeding, provisioning, storage) and ecological or physiological constraints on time and energy (*5*, *11*, *12*). Different groups of organisms maximize currencies of efficiency or rate in accordance with these principles (*11*–*15*). The legacy of metabolic acceleration in humans strongly suggests that our unique subsistence strategies are characterized by novel ways to mitigate time and energy constraints in pursuit of high-value foods that expand daily caloric acquisition beyond that of other primates.

There is substantial evidence to suggest that humans expend less energy and time on subsistence than other great apes. Humans exhibit several derived features that reduce energetic costs, including anatomical and behavioral traits that reduce the cost of walking and searching (*16*) and reductions in the size of costly digestive organs (*17*). Compared to other primates, humans also spend far less time feeding (*18*), use more sophisticated tools to acquire and process foods, and engage in cognitively complex and hypercooperative behaviors to obtain energy-dense foods that would otherwise be inaccessible (*2*). Yet it remains unknown to what extent these derived attributes actually reduce the energy and/or time costs of subsistence, thus requiring direct comparisons of the costs and benefits of food acquisition between humans and other great apes.

To clarify how unique human subsistence strategies enabled metabolic acceleration and surplus production for provisioning, we investigated whether humans achieve greater efficiency or acquisition rate relative to other great apes. We calculated subsistence costs (energy and time) and energy acquisition among wild orangutans, gorillas, and chimpanzees, and compared these measures with high-resolution data collected among Hadza hunter-gatherers in Tanzania and Tsimane forager-horticulturalists in Bolivia (Fig. 1). Both populations actively forage (hunt, gather), while the Tsimane also practice slash-and-burn horticulture, which permits exploration of further changes in the energetics of subsistence associated with farming. Whereas many studies on the ecological economics of humans are concerned primarily with the "exosomatic metabolism" that distinguishes humans from other organisms (i.e., energy flows metabolized outside of the body, e.g., from wood fuel, domestic animals, industrial processes), our focus here is on "endosomatic metabolism" (i.e., food energy used within the body) (*19*, *20*). Small-scale subsistence societies such as the Hadza and Tsimane use far more exosomatic energy than other organisms in the form of wood fuel for cooking or land management; tools for hunting, digging, food processing, or field preparation; and the occasional use of dogs for hunting (*21*–*23*). Here, however, we

[1]Department of Anthropology, University of California, Santa Barbara, CA, USA. [2]Department of Human Behavior, Ecology, and Culture, Max Planck Institute for Evolutionary Anthropology, Leipzig, Germany. [3]Department of Anthropology, University of Utah, Salt Lake City, UT, USA. [4]Institute for Advanced Study in Toulouse, Toulouse, France. [5]Department of Anthropology and Archaeology, University of Calgary, Calgary, Alberta, Canada. [6]Department of Anthropology, University of New Mexico, Albuquerque, NM, USA. [7]Department of Anthropology, University of Nevada, Las Vegas, NV, USA. [8]Department of Anthropology, University at Buffalo, Buffalo, NY, USA. [9]Department of Anthropology, University of California, Los Angeles, CA, USA. [10]Institute of Human Origins, School of Human Evolution and Social Change, Arizona State University, Tempe, AZ, USA. [11]Human and Evolutionary Biology Section, Department of Biological Sciences, University of Southern California, Los Angeles, CA, USA. [12]Department of Evolutionary Anthropology, Duke University, Durham, NC, USA. [13]Duke Global Health Institute, Duke University, Durham, NC, USA.
***Corresponding author. Email: tkraft@anth.ucsb.edu (T.S.K.); herman.pontzer@duke.edu (H.P.)**
†These authors contributed equally to this work.

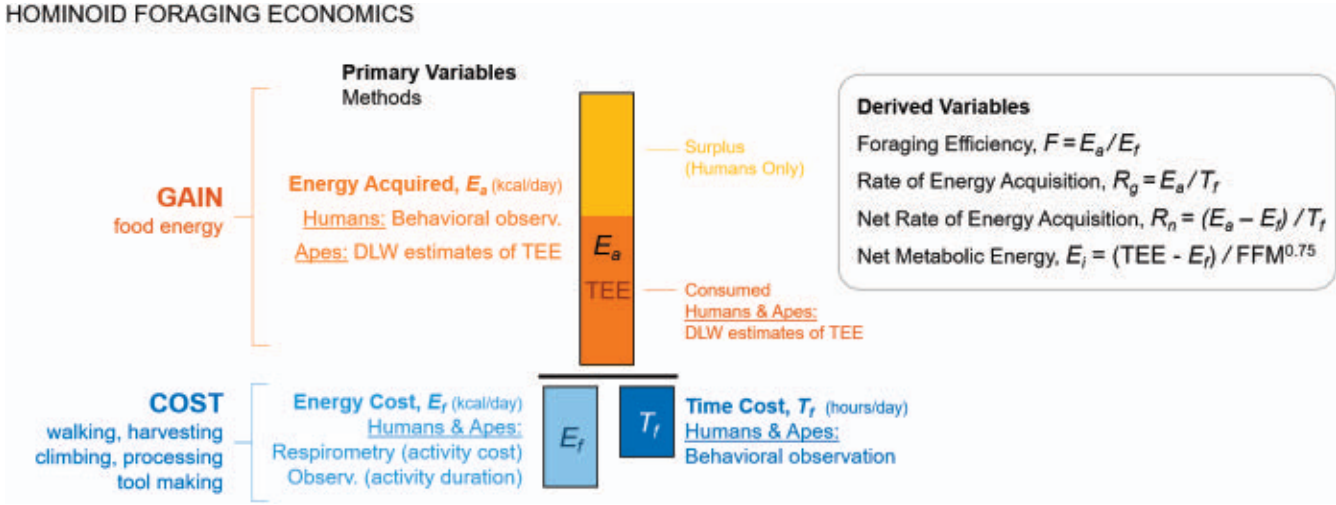


Fig. 1. Overview of methods and variables used to compare foraging economics in humans and other great apes. Energy acquired, E_a, for humans was determined through behavioral observation of food production. Humans consume some of the energy they acquire (equal to their total energy expenditure, TEE) and share or store remaining surplus. For other great apes, food sharing and provisioning are negligible (*2*, *29*–*31*), and therefore E_a = TEE. In humans and other apes, TEE was estimated from DLW measurements. The energy cost of foraging for humans and other apes, E_f, was calculated by multiplying the energy costs of foraging activities (measured using respirometry) by the time spent in each activity (determined from behavioral observation). The time cost of foraging for humans and other apes, T_f, was determined from behavioral observation. These primary variables (E_a, E_f, and T_f) were used to derive foraging efficiency (F), rate of energy acquisition (R_g), net rate of energy acquisition (R_n), and net metabolic energy (E_i). E_i is the energy consumed and available for nonforaging tasks, scaled to metabolic body size, $FFM^{0.75}$, where FFM is fat-free mass.

limit our consideration of time and energy costs to endosomatic energy flows that are critical for understanding how biological tradeoffs and constraints in energy budgets affect the ability to support expensive organs and life history traits. We therefore define system boundaries in this study to include only those behaviors that directly or indirectly contribute to tasks relating to food acquisition, processing, and consumption. We included both foraging and horticultural activities in subsistence cost estimates, as well as auxiliary behaviors such as food processing, tool manufacture, eating, and firewood/water collection. Energetic costs of subsistence tasks were calculated as net values (i.e., resting costs were subtracted) to parse out the additional costs of activities above baseline.

For the Hadza and Tsimane, we used long-term observational data on food acquisition and production to establish population-average rates of daily energy acquisition, E_a (kcal/day) for men and women, and we measured rates of energy expenditure in subsistence activities using a portable respirometry system in the field (Fig. 1). We then integrated time allocation data (scan sampling and focal follows) with respirometry-based measures of the energetic costs of subsistence tasks to estimate the following: daily energetic cost of subsistence, E_f (kcal/day); daily time spent on subsistence, T_f (hours/day); efficiency of subsistence [$F = E_a/E_f$; note that F is equivalent to the modified form of efficiency in (*12*) because foraging activities are measured net of resting costs]; and gross ($R_g = E_a/T_f$) and net [$R_n = (E_a - E_f)/T_f$] rate of energy acquisition (kcal/hour). Finally, we calculate a quantity, E_i [$= (TEE - E_f)/FFM^{0.75}$, where TEE = total daily energy expenditure in kcal/day measured using doubly labeled water (DLW), and FFM = fat-free mass], which represents the net energy available to the body for nonsubsistence purposes scaled by metabolic body mass (*24*).

We used a similar approach to calculate these variables for nonhuman great apes. Using published DLW measures of TEE for zoo-living chimpanzees, gorillas, and orangutans, we fit species-specific regressions of TEE against body mass and used them to estimate TEE for adult males and females of each species in the wild (see methods). The DLW method is considered the gold standard for measuring TEE in free-living conditions (*25*), and our DLW-based estimates of TEE for wild apes were similar to TEE estimates based on activity budget analyses in wild apes and other primates (*26*) and to DLW measurements of TEE in populations of wild primates (*27*) and other mammals (*28*) (table S1 and fig. S1). Further, the TEE–body mass relationship is similar for captive and wild primates (*27*), justifying the use of data from zoo-living animals in regressions to estimate TEE in wild apes. Nonetheless, to test the sensitivity of our results to the use of E_a estimates derived from TEE, we also used estimates of daily caloric consumption from feeding observation studies in wild apes, which are generally higher than our DLW estimates, for comparison (table S1). Results from those analyses did not substantially affect the pattern of differences between other ape species, nor between other apes and humans (see methods).

Among nonhuman great apes, average TEE = E_a because provisioning and food storage are negligible (*2*, *29*–*31*). Time spent foraging and distances traveled and climbed per day were compiled from observational studies of wild apes. These data were then used to calculate the remaining variables using locomotor costs from published respirometry studies of energy expenditure during walking in chimpanzees (*32*) and climbing in other primates (*33*) (see methods).

We tested the hypothesis that human subsistence strategies reduce both energy (E_f) and time (T_f) costs of food acquisition and increase gross energy acquisition (E_a) relative to the strategies of other great apes, thus improving the energetic efficiency of subsistence (F), gross or net energy return rates (R_g, R_n), and net energy availability (E_i). Given that agriculturalists generally have higher fertility rates than hunter-gatherers (*34*, *35*), we predicted that Tsimane horticulturalists would evince greater daily energy acquisition, efficiency, and return rates than Hadza hunter-gatherers despite potentially expending more time and energy on subsistence. Finally, to test whether the results obtained from the Hadza and Tsimane are representative of hunter-gatherers and horticulturalists more broadly, we assembled a database on efficiency, production, energy costs, and time allocation during subsistence from a global sample of contemporary hunter-gatherer (n = 14) and horticulturalist populations (n = 22) (table S2). This database allowed us to address

debates over fundamental differences in the energetics of hunting and gathering versus farming (*36*, *37*). All results were analyzed separately by sex to discern how the sexual division of labor characteristic of human societies affects the distribution of time and energy costs.

Subsistence energetics for humans versus other great apes

Contrary to our expectations, we found that adult human hunter-gatherers and horticulturalists exhibited higher subsistence energy costs (E_f) than other great apes (Fig. 2 and table S2). E_f for humans [mean (95% CI), men: Hadza 709 (638, 774) kcal/day; Tsimane 603 (403, 962) kcal/day; women: Hadza 474 (339, 629) kcal/day; Tsimane 253 (205, 320) kcal/day] was substantially greater than for other great apes (males: 152 to 293 kcal/day; females: 76 to 168 kcal/day). Consequently, whereas gross energy acquisition (E_a) was greater for humans, subsistence efficiency (F) was similar between humans (range: ~5 to 16) and other great apes (range: ~8 to 16) (Fig. 3). Females in all species had lower energetic costs of subsistence (E_f) than males (Fig. 2), increasing energy available for reproduction (*38*). The energetic efficiency of Tsimane subsistence was higher than estimates for other nonindustrial societies obtained from the literature (Fig. 3, fig. S2, and table S2) but similar to crop-specific efficiencies for tropical horticulturalists reported elsewhere [F = 10 to 20, not accounting for transport and processing costs (*39*)]. Notably, subsistence efficiencies for humans and other great apes are at the low end of the observed range for vertebrates, most of which exhibit F > 40 (*40*).

Humans devoted considerably less time (T_f) to subsistence than other great apes, who spend 7 to 8 hours/day on subsistence (Fig. 2). Differences in time spent on subsistence between humans and other great apes would be even greater if calculated on the basis of a more limited range of activities (e.g., excluding time spent on tool manufacture) or as a percentage of all waking hours, as humans sleep less than any other primate (*41*). Accordingly, both gross (R_g) and net (R_n) energy acquisition rates were substantially higher for humans than for other great apes (Fig. 3). Although energy acquisition rates for other great apes were somewhat higher when using observational data on wild food intake to estimate E_a (as opposed to DLW-based regressions for TEE), the same general pattern was observed (fig. S3). A decomposition of net energy acquisition rates reveals that these species differences were driven by reductions in both time spent on subsistence and increases in gross energy acquisition, but not reduced costs. Indeed, humans expended more energy on subsistence (E_f) than any other great apes (Fig. 2).

The fact that humans expend more energy on subsistence than other great apes, despite spending less time, indicates that humans engage in comparatively high-intensity foraging activities. Whereas other great apes expend little energy during foraging, humans engage in a diversity of specialized and energetically costly tasks with the use of tools. Hadza and Tsimane chop, dig, climb, burn, manufacture objects, and transport food, fuel, and water to central places (Fig. 4). Tsimane additionally invest in strenuous activities associated with farming, such as clearing brush, felling trees, and transporting crops. An additional database we compiled on the instantaneous energy costs (kcal/min) of common activities in nonindustrial societies reveals

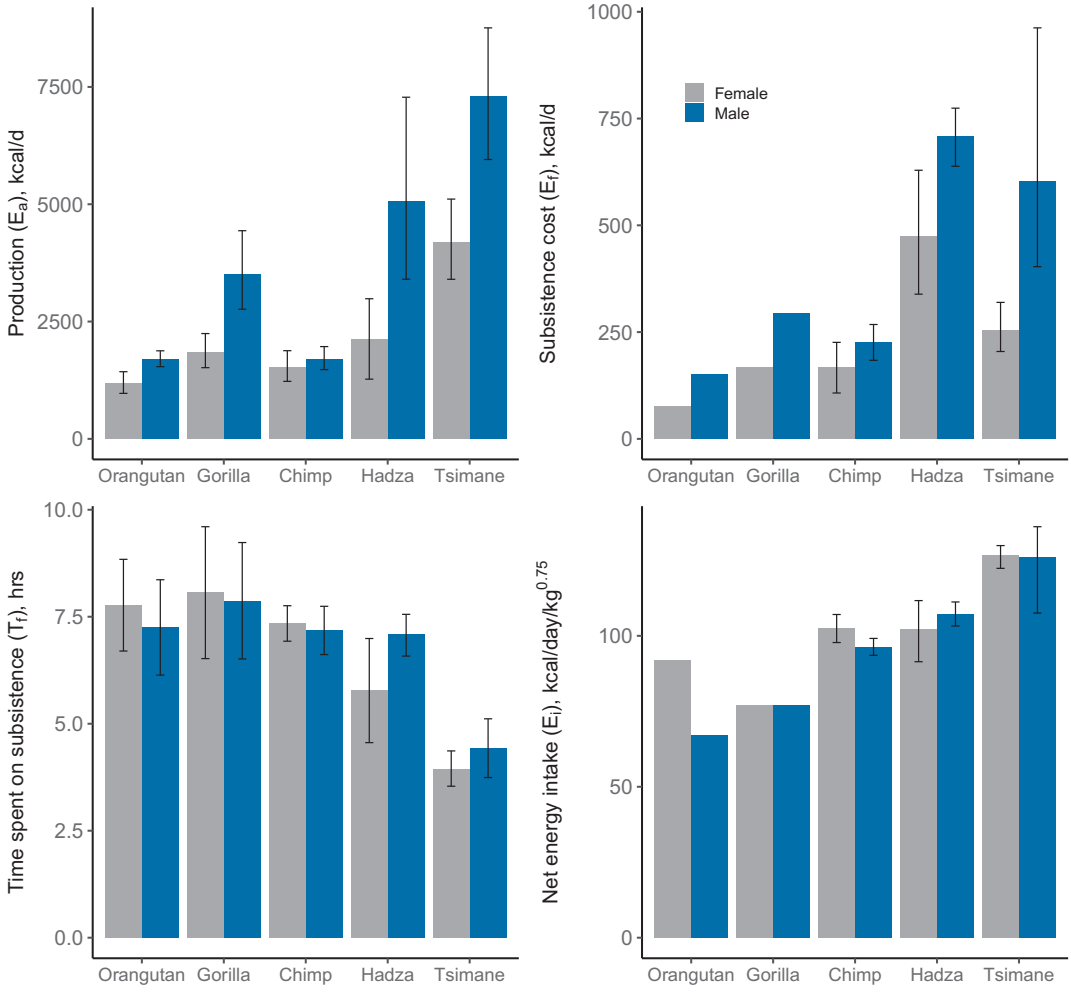


Fig. 2. Cross-species comparison of daily energy acquisition (E_a), subsistence costs (E_f), time spent on subsistence (T_f), and net energy intake (E_i). Values represent posterior medians ± 95% HDPIs (for humans) or means ± 95% CIs (for other great apes). Note that some values are missing error bars because only point metrics were available for cost calculations.

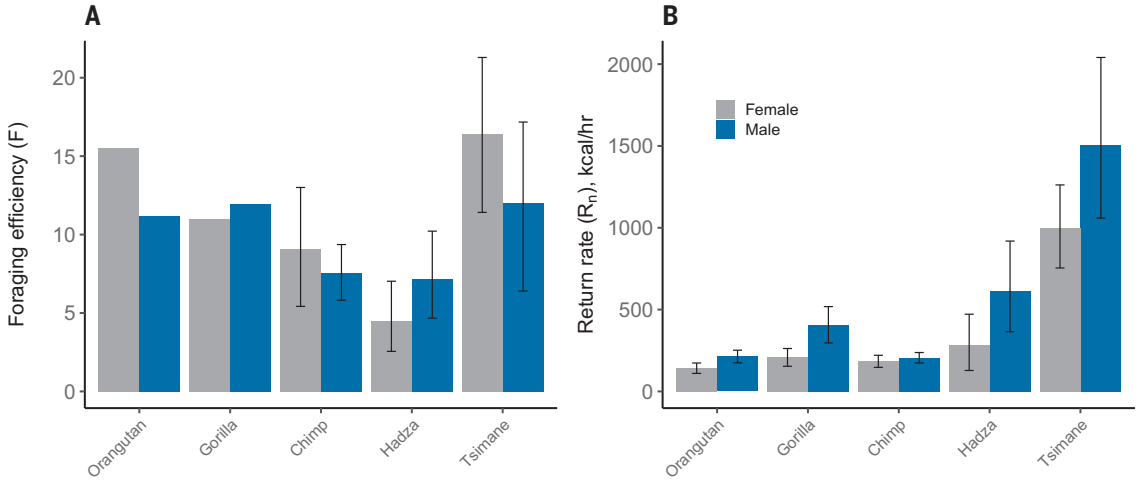


Fig. 3. Energy spent on subsistence across species/populations. (**A**) Energy acquired per unit energy; (**B**) energy acquired per unit time. Values represent posterior medians ± 95% HDPI (Tsimane/Hadza) or means ± 95% CI (all others). Tsimane and Hadza estimates represent population estimates for a 40-year-old individual.

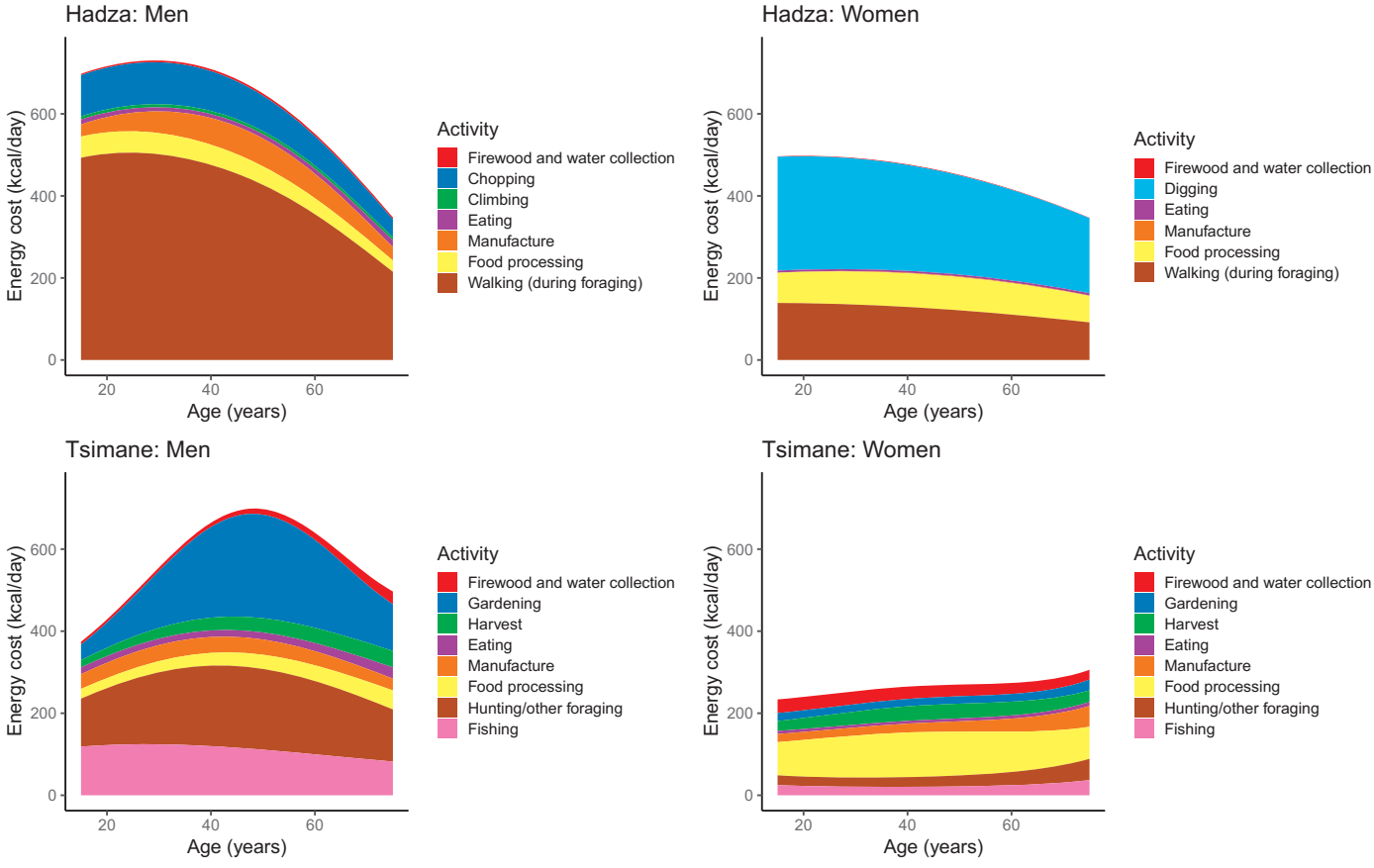


Fig. 4. Subsistence costs for Hadza and Tsimane over the life course. Costs (kcal/day) reflect the time that individuals of different ages spend on activities as well as the relative energetic costs of engaging in those activities.

that many human subsistence activities are energetically costly, and that tasks associated with farm work bear marginally higher costs than those associated with foraging (including walking) (Fig. 5).

Net energy availability (E_i) scaled by fat-free body mass (*1*, *24*) was generally greater among humans than other great apes, especially for Tsimane (Fig. 2). Human subsistence, and particularly strategies including farming, thus yields more energy that can be used by the body for diverse purposes other than subsistence, even though humans spend a greater absolute proportion of their TEE on subsistence (Fig. 6). This greater net energy availability likely enables

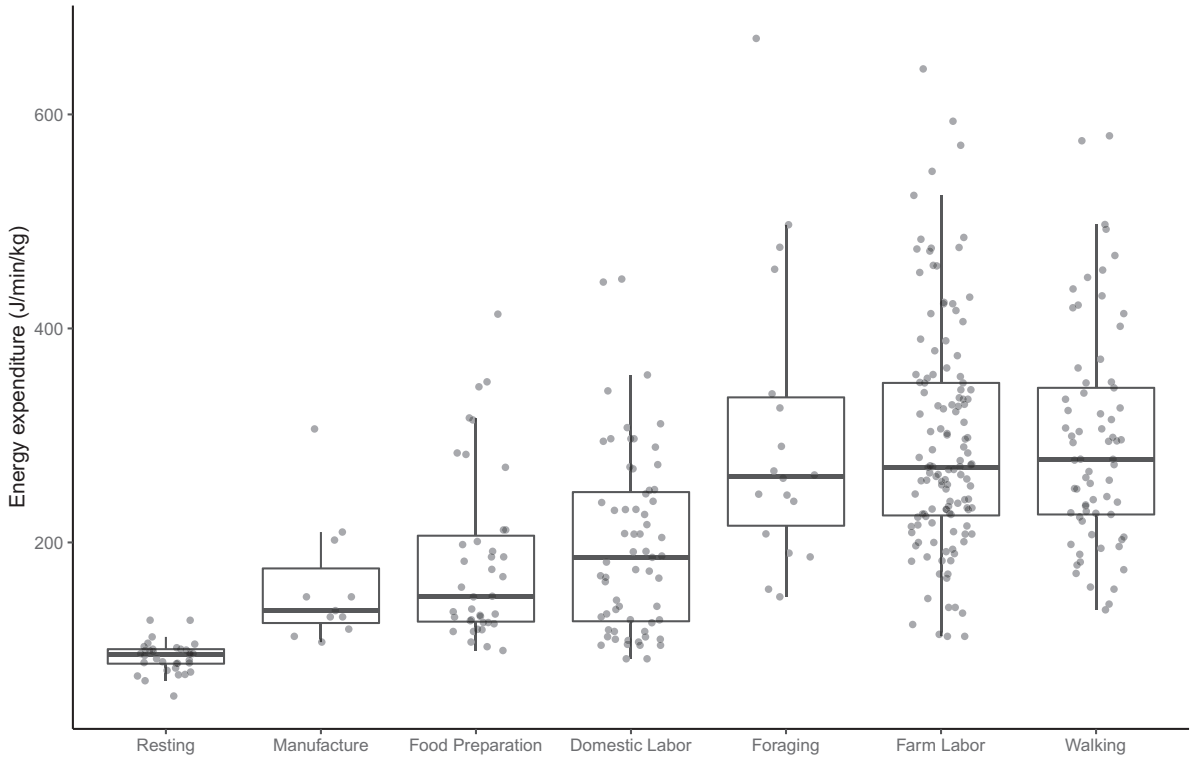


Fig. 5. The instantaneous metabolic cost of activities undertaken in traditional subsistence societies. Values were either measured empirically for this study or derived from the literature (table S5). Data from field activities are measured using respirometry. Many foraging activities involve walking but were categorized under foraging according to their original designations in the literature. Costs are presented for both men's and women's activities and are standardized per kilogram of body mass.

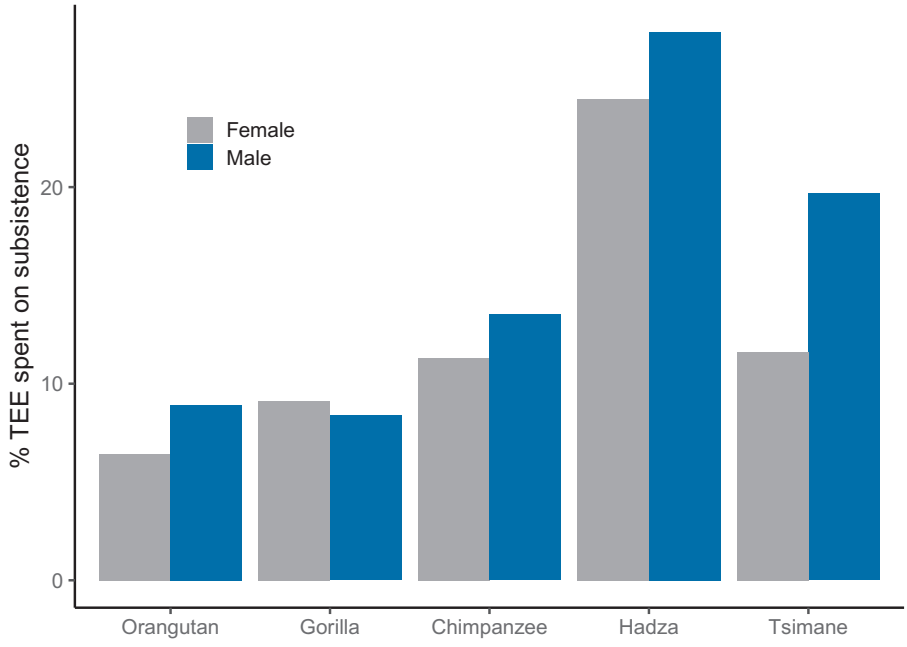


Fig. 6. The percentage of total energy expenditure (TEE, kcal/day) spent on subsistence costs across species/populations.

humans to afford the high maintenance costs of large brains and elevated reproductive investment (*1*), although the substantial energy costs of human subsistence limit interspecific differences in E_i and suggest that the primary energetic benefit of human subsistence strategies is obtaining surplus energy for provisioning children, reproductive females, and other group members. Unlike other great apes, humans exhibit extensive surplus production by adults (E_a > TEE), which is critical for provisioning females bearing the energetic costs of reproduction (*42*, *43*) and young offspring that are net consumers (*2*), thus enabling elevated fertility despite compound offspring dependency (*44*, *45*). Provisioned children and reproductive-aged women also have lower subsistence costs (*46*) and thus have more energy available during critical periods of brain growth and development (*47*) and gestation and lactation, respectively.

Our results indicate that horticulture may improve efficiency (F) relative to hunting and gathering, especially for females (Fig. 3). Horticulture also reduces time costs and increases daily gross energy acquisition, leading to substantially higher daily and hourly return rates (by a factor of ~1.5 to 2; Figs. 2 and 3). These differences persist across ontogenetic development (Fig. 7). Within-population comparisons further demonstrate that activity-specific energetic return rates from horticulture exceed those of hunting and gathering in mixed economies (table S3). These results are consistent with previous comparisons between Amazonian horticulturalists and hunter-gatherers (*48*–*52*). The potential for elevated energetic return rates could therefore have favored the adoption of incipient agriculture, particularly in environments suffering from depletion of profitable wild food resources (*37*).

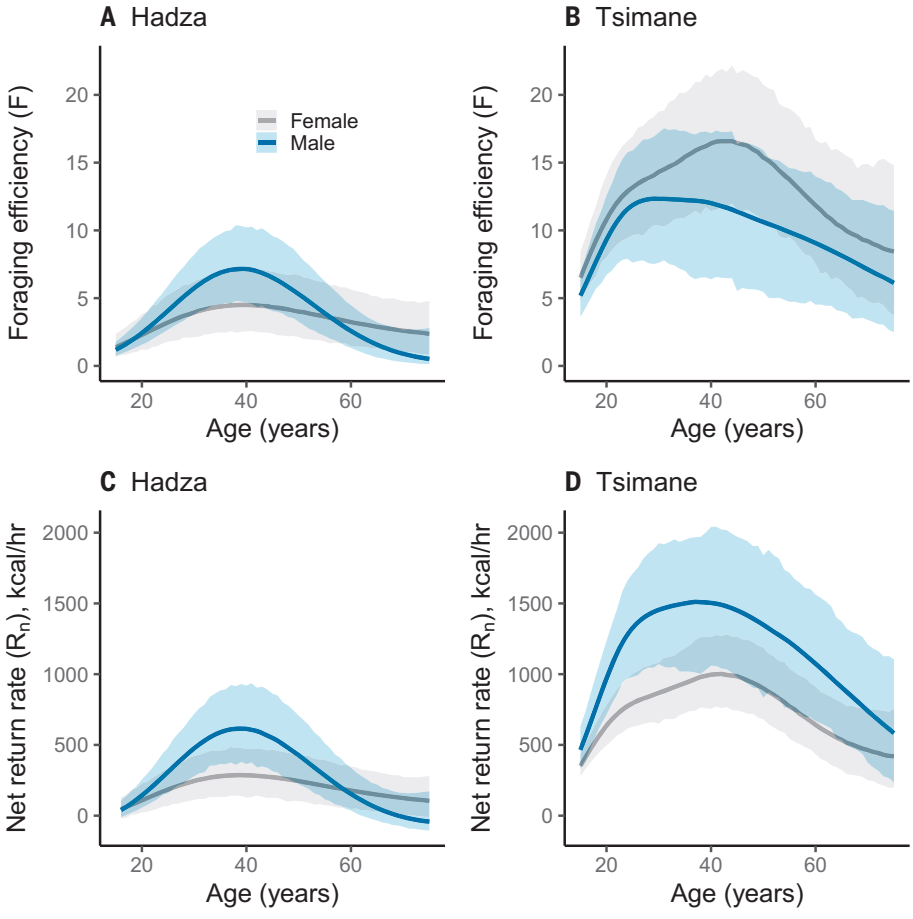


Fig. 7. Efficiency (*F*) and net return rates (R_n) across the life course in Hadza and Tsimane. (**A** and **C**) Hadza *F* and R_n; (**B** and **D**) Tsimane *F* and R_n. Data are separated by sex (blue = male, gray = female). Lines and shaded regions represent predicted mean values and 95% CIs based on composite calculations. Tsimane exhibit greater net return rates than Hadza across all ages (in years).

Cross-cultural energetics of human hunter-gatherers and horticulturalists

Our cross-cultural database of traditional hunter-gatherers and horticulturalists ($n_{\text{hunter-gatherer}}$ = 14, $n_{\text{horticulturalist}}$ = 22) confirms our findings from detailed study of the Hadza and Tsimane (table S2). Specifically, we found further evidence that humans of both subsistence modes produce more calories per day, spend less time on subsistence, and have higher return rates than other great apes (Fig. 8; see fig. S4 for results using observational acquisition data for nonhuman great apes). The few available published estimates of efficiency in subsistence horticulturalists also support the observation that human efficiencies are not elevated above those of other great apes (fig. S2). Among humans, cross-cultural analyses also show that horticulturalists produce more daily calories [combined sexes for all societies; mean (kcal/day) = 7520 versus 3061, $t(10.3) = 2.4$, $P = 0.048$], spend similar amounts of time on subsistence [mean (hours/day) = 4.1 versus 4.5, $t(21.4) = -0.7$, $P = 0.48$], and have higher return rates [mean (kcal/hour) = 2162 versus 729, $t(9.2) = 2.2$, $P = 0.052$] than hunter-gatherers, in general agreement with results for the Hadza and Tsimane. Sex differences in the cross-cultural sample were minimal for the amount of time spent on subsistence and return rates (all $P > 0.2$), but among hunter-gatherers, daily production was higher in men than in women [mean (kcal/day) = 3879 versus 2243, $t(17.9) = 3.3$, $P = 0.004$].

Despite fairly large average differences between hunter-gatherers and horticulturalists, cross-cultural estimates indicate a great deal of variability within subsistence mode for all variables assessed (Fig. 8). For example, the amount of time spent on subsistence ranges from 2.6 to 7.0 hours/day for hunter-gatherers and from 1.8 to 8.5 hours/day for horticulturalists (pooled sexes). Likewise, although return rates are generally higher for horticulturalists than hunter-gatherers, the distributions of outcomes overlap between subsistence modes. This suggests that hunting and gathering can be equally or even more profitable than farming, depending on the local environment, and may help to explain overlap in total fertility across subsistence modes (*35*) as well as flexibility and reversions in subsistence patterns (*37*).

Energetics and the origins of hunting and gathering

Overall, our results point to a revised understanding of the evolution of hunting and gathering. Whereas complex technology and behavioral innovations undoubtedly reduce the energy costs of particular subsistence tasks (*53*–*59*), they are used in pursuit of resources that are energetically costly to acquire (e.g., meat, tubers, baobab) and necessitate greater day ranges than those of other primates (Fig. 4) (*26*, *60*). Humans therefore accommodate the energetic demands of metabolic acceleration and surplus energy acquisition for provisioning not by decreasing subsistence costs and increasing energetic efficiency, but instead by expending larger amounts of energy to rapidly attain larger amounts of energy from the environment (Figs. 2 and 3). The emergence of hunting and gathering could have been favored despite increased energy costs, as long as they increased net acquisition rates and total energy gained (*9*–*11*, *61*). Thus, hunting and gathering likely evolved with a focus on high-quality dietary resources made available by extractive foraging and complex processing (*2*) in combination with behaviors and technologies leading to the minimization of time spent foraging (T_f) at the expense of increased energy (E_f) costs.

The finding that humans are not particularly efficient foragers runs counter to a prevailing assumption in the anthropological literature, which presumes that the suite of human adaptations (e.g., complex tool use, bipedalism) and our unique ability to harness exosomatic energy have served to reduce endosomatic expenditure (*62*, *63*). However, the results of this study are consistent with the predictions of optimal foraging theory. Human subsistence is best contextualized using foraging models of provisioning, in which an organism seeks to maximize total daily energy delivery subject to meeting its own energetic requirements (*11*). In the absence of constraints, an optimal forager maximizes the net rate of energy gain (R_n). Under energetic or time constraints, however, a forager may do better to maximize efficiency, particularly when limited by the amount of energy the forager can assimilate/expend or a low self-feeding rate (rate at which food is acquired to self-provision) (*5*, *11*, *64*).

Human foraging exhibits characteristics that promote optimal rate-maximizing behavior under either energy or time constraints. Metabolic acceleration (including increased TEE), coupled with economical bipedal locomotion, large fat stores, and enhanced heat dissipation capacity, increases the ceiling on the maximum daily energy expenditure of a forager while simultaneously decreasing the rate at which that

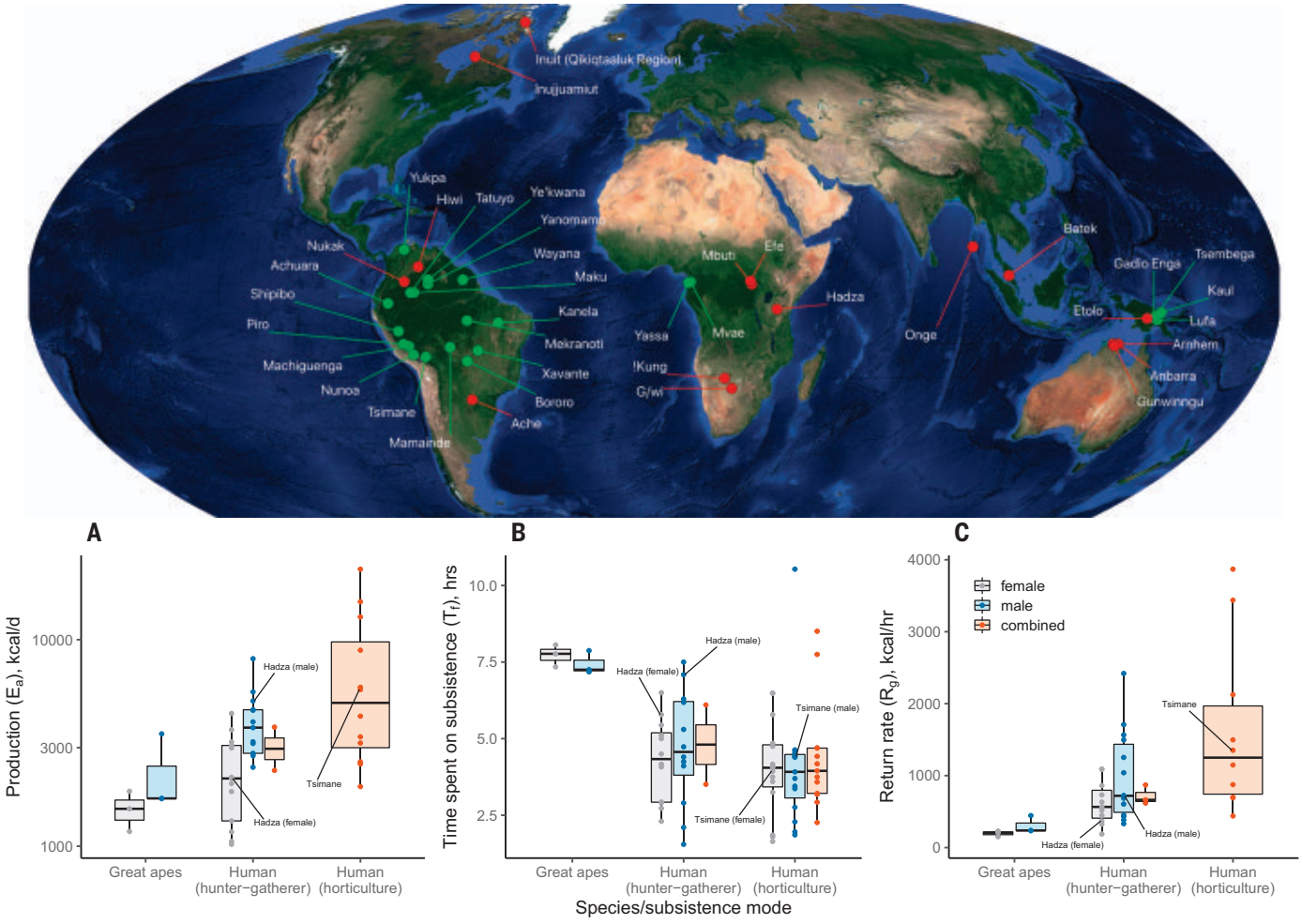


Fig. 8. Comparison of nonhuman great apes, hunter-gatherers, and horticulturalists using the cross-cultural sample. World map shows the location of societies (red = hunter-gatherer, green = forager-horticulturalist) included in the cross-cultural sample. In (**A**) to (**C**), novel values for Tsimane and Hadza from the current study are included and labeled for reference. Male and female values for Tsimane were averaged for measures of production (E_a) and gross return rate (R_g) because no separate values were available from the literature by sex. Note that the *y* axis is on a $\log_{10}$ scale in (A). In (C), one high outlier is omitted for "Human (horticulture)" for visualization purposes.

ceiling is approached (*1*). Humans also have a number of traits that increase the self-feeding rate, which is critical for supporting high-intensity foraging (*11*). Whereas other primates eat raw foods that require long digestion times, humans target energy-dense foods (*65*) that become highly digestible with processing (*66*, *67*) and produce large surpluses that can be pooled, stored, and distributed. A human forager may effectively experience less self-provisioning risk because of the expectation that other family or camp members will bring surplus resources back to a central place for later consumption, thereby enabling rate-maximizing foraging strategies that would otherwise be too risky (i.e., high potential gain, but also high potential for failure) (*11*). In addition, hunter-gatherers are known to consume easily digestible, high-energy foods out of camp while foraging, which can rapidly provide energy to sustain further food acquisition; our data indicate that Hadza men and women consume approximately 80% and 20% of daily TEE while out of camp, respectively (*68*).

High-intensity foraging strategies are not unique to humans and have arisen numerous times throughout the tree of life. The best examples come from social insects. For instance, ants exhibit great variation in the "tempo" of foraging, in which high foraging intensity correlates with increased resource abundance, colony size, and caste polymorphism (*69*), or the availability of excess carbohydrate resources from the exudates of aphids that may enable high-energy activities without compromising colony growth (*70*). Intriguingly, farming (of fungal cultivars) has emerged in hundreds of ant species (*71*), and it would be interesting to determine whether the energetic advantages obtained mirror the apparent impact of the domestication of plants and animals by humans (e.g., increased population growth rates). Asian honey bee species also show marked differences in mass-specific metabolic rates and the rate of performance of colony tasks, a pattern that appears to be linked with the demands of provisioning and defense of open versus cavity nests (*72*). More generally, parental care may have driven the remarkable convergent evolution of endothermy and the associated ability to sustain vigorous exercise among birds and mammals (*73*).

High-intensity foraging and cooperative energetics

A high-throughput foraging strategy in humans coevolved with cooperative provisioning and alloparental care to produce our characteristic life history strategy. In the human case, this shift occurred within the context of an existing scaffolding of cognitive abilities, incipient tool use, advanced food processing, and extractive foraging (*2*, *74*). The resulting energy system was in a sense risky and was made possible only by lifting the foraging time constraints that characterize other great apes while simultaneously increasing tolerance for high

daily energetic costs. Our findings thus help to unify and shed light on two dominant theories in human evolutionary studies: (i) the embodied capital model, which attributes large brain growth and a long developmental period in humans to the need to learn and develop difficult skills associated with extractive food acquisition (*2*), and (ii) the cooperative breeding hypothesis or pooled energy budget model, which posits that many derived features of human behavior, intelligence, and cognition are linked to a cooperative provisioning system that arose after the evolutionary split with the *Pan* lineage (*38*, *75*, *76*). Our findings suggest that increased daily production associated with extractive foraging in humans (*2*) was enabled by increased foraging intensity that reduced the time cost, but not the energy cost, of food acquisition. Such a strategy is unlikely to be tenable in the absence of a cooperative production and provisioning system with widespread sharing, divisions of labor, alloparental care, and prosociality. As we have shown, human subsistence strategies demand a large proportion of TEE (Fig. 6) and (at least in hunter-gatherers) the proportion of days in which individuals fail to produce food can be high (fig. S5). Humans are thus more prone to fatigue and starvation than other great apes when expected returns fail to materialize, especially given that humans often pursue high-risk/high-reward foods, such as large game (*77*). Some high-intensity activities (e.g., climbing or running) may also put foragers at an elevated risk of injury. Age-related increases are evident in the energetic efficiency of subsistence and in rates of energy acquisition among humans (Fig. 7), with intergenerational transfers of surplus calories by individuals at ages of peak productivity buffering risks of energetic shortfalls. These transfers, along with divisions of labor, the capacity for storage, and pooled energy budgets (*38*), allow for wide distribution of surplus resources to buffer adults as well as offspring. The pooled energy budgets and food storage capabilities of humans thus render it profitable to pay high energy costs and pursue high-risk and/or slow-to-master foraging behaviors in order to capitalize on lucrative return rates.

We argue that the extraordinary energy surpluses produced by adult humans, and our expensive encephalized brains, are unattainable for an organism with a foraging strategy and locomotor economy like those of other great apes. In order to achieve the same proportional surplus of a 40-year-old Hadza male (99% increase above TEE), for example, an adult male chimpanzee would need to forage ~14.3 hours/day; an adult female chimpanzee would have to forage ~14.0 hours/day to achieve the same surplus (91% increase above TEE) of a 40-year-old Tsimane woman. Human day ranges would also introduce prohibitive locomotor costs if terrestrial locomotion were as uneconomical as that of a chimpanzee: For a 50-kg human moving the same daily distance traveled by male Hadza [~14 km (*78*)] with the locomotor economy of a bipedal chimpanzee (1.06 kcal kg^{-1} km^{-1}), daily travel costs alone would be ~750 kcal/day, which would be ~30% of TEE, and nearly 3 times the cost of traveling that distance with a standard human locomotor economy (~270 kcal/day; 0.39 kcal kg^{-1} km^{-1}).

Superficially, these calculations seem to support the argument that the evolution of greater walking economy helped to facilitate encephalization in human evolution by reducing travel costs and thereby making more energy available for an increasingly expensive brain (*62*, *63*, *79*). In this scenario, the reduced cost of bipedal walking in early members of the genus *Homo* would have enhanced foraging efficiency, helping to pay for a larger brain. However, enhanced walking economy would not by itself save much energy without a change in subsistence strategy. For example, imagining a 50-kg male hominin with fully human-like walking economy but retaining chimpanzee-like foraging and daily travel distances [~2.5 km/day (*80*)], we estimate that this individual would have saved only ~80 kcal/day compared to the costs for a chimpanzee traveling the same distance. This saving, while not trivial, likely could not have improved foraging efficiency enough to fuel the increased encephalization that occurred with the appearance and evolution of *Homo*. However, lower walking costs in bipedal hominins could have rendered the longer travel distances required in hunting and gathering to be more energetically feasible, ultimately allowing early members of the genus *Homo* to adopt a subsistence strategy dependent on long day ranges. High-intensity foraging activities, in addition to greater day ranges, would have favored the evolution of a high-endurance phenotype in humans relative to other apes (*81*–*83*). In this scenario, the great energy gains needed to fuel increasing hominin brain sizes would have been achieved not through greater foraging efficiency (*F*), which remained low and similar to other great apes, but instead through the ability to engage in hunting and gathering strategies that required high daily travel costs but yielded very high energy rewards at a fast rate.

Thermal (exosomatic input from fire) and nonthermal (e.g., pounding, winnowing, fermenting) food processing would likewise have helped to enable high-intensity foraging strategies. Processing, the intentional external modification of a food resource to alter its physical and/or chemical attributes in preparation for consumption (*22*), can improve energy capture in several important ways. Even minimal food processing can effectively increase digestibility and bioaccessibility, reduce pathogens, and denature toxins (*67*, *84*, *85*). Thermal processing, particularly cooking, not only alters the nutritional quality of foods, but also begins the externalized phase of digestion (*84*, *86*). Cooking can substantially reduce the costs of meat digestion, absorption, and assimilation (*87*) and reduces the physical structure of plants (starch, inulin, cellulose), increasing digestibility of their basic nutritional elements. Nonthermal processing is also likely a key component of high-intensity foraging and works in many of the same ways, with the addition of particle size reduction prior to ingestion (*22*). Our findings suggest that although some food processing activities can be energetically intensive (Fig. 5), overall amounts of time and endosomatic energy devoted to processing remain relatively low (Fig. 4). Given the ability to increase edibility and digestibility of foods, processing should thus yield a high efficiency or return on investment. Although our analyses do make use of caloric estimates of food items after undergoing thermal and nonthermal processing, a limitation of the current study is that our measures of subsistence costs do not include the energetic cost of digestion and thus any potential cost savings that humans experience digesting processed foods.

A central finding here is that humans devote less time to subsistence activities than other great apes (Figs. 2 and 8). Humans are also the only primate species that can afford to take rest days (days in which individuals voluntarily choose not to forage) on account of our reliance on cooperation, sharing, and pooled energy budgets. With less time spent foraging, ancestral hominins would have experienced greater opportunities for alternative activities (*10*, *88*, *89*), including cultural production and exchange, by loosening constraints on the demanding foraging time costs observed in other great apes. Through improvements to technology and social exchange of information, such investments would have enabled further increases in energy acquisition rates, in turn freeing more time to invest in nonsubsistence pursuits. These include quintessential human behaviors that occur in the context of central-place foraging—social learning, object manufacture, and symbolic/ritual activities—which, combined with social tolerance and bilateral networks, could collectively favor ongoing processes of cumulative cultural evolution (*90*, *91*).

Energetics and the origins of farming

Energetic and time considerations lie at the heart of theories to explain the origin and spread of agriculture. In particular, it has been hypothesized that the adoption of farming could have been a response to a higher marginal return on labor (*36*, *37*). Using high-resolution measurements of subsistence energetics from the Hadza and Tsimane in combination with a cross-cultural sample, our results indicate that horticulture is generally associated with higher

return rates compared to hunting and gathering, despite there being little difference in the amount of time devoted to subsistence (Fig. 8). As such, higher return rates arise primarily from greater caloric production within the same amount of time. These results support prior evidence that the adoption of farming could have been motivated by economic factors, primarily greater gains per amount of work (time) spent on labor (*37*). Technological improvements such as the shift from stone or wooden tools to metal tools may decrease the time required for agricultural work and improve return rates (*55*, *59*). Estimated return rates and efficiencies for contemporary subsistence populations may therefore be higher than they would have been for early humans lacking modern technology (e.g., machetes), although some modern environments may also be more depleted (*92*).

Our results further contradict any notion of the "original affluent society," according to which hunter-gatherers work ~15 hours/week (*93*, *94*): Hadza men and women work ~50 and 40 hours/week and other hunter-gatherer men and women work ~33 and 28 hours/week on average, respectively, based on our cross-cultural sample (table S2). Agriculture compared to hunting and gathering is also not necessarily accompanied by increased working time, as has been hotly debated (*95*). Although some farming groups do work more hours, on average there is no difference in total work time between modes of subsistence (Fig. 8). The wide range of times devoted to subsistence among both hunter-gatherers and horticulturalists suggests that local ecological and social factors, rather than subsistence mode, dictate available leisure time.

Our results also provide a proximate mechanism to explain the elevated reproductive rates often associated with the shift to agriculture (*4*, *96*), which has been linked to increased available energy for women to invest in reproduction. Tsimane women expend 47% less energy on subsistence than Hadza women (Fig. 2), devote less than half the amount of TEE to subsistence (Fig. 6), and have more energy available for reproduction scaled to body mass (E_i; 18% greater). Reductions in energy expenditure may result from Tsimane women engaging primarily in tasks that require little energy, such as food processing and domestic labor (Fig. 5), in contrast to Hadza women, who spend nearly 40% of time out of camp engaged in intensive digging for underground plant foods (table S4). The relative subsistence costs for Hadza and Tsimane women correspond to observed differences in total fertility rates [TFR_{Hadza} = 6.2 (*21*), $TFR_{Tsimane}$ = 9.1 (*97*)].

Recent changes in human energetics

We have shown that human subsistence has evolved to capture ever-greater amounts of surplus energy quickly, but at the expense of high energy costs. Such high costs persist even though humans have economical locomotion (the primary energetic cost of foraging for most animals) and use tools that reduce the costs of particular foraging activities relative to the cost of the same activity performed without tools. This suggests that energy gained from improvements to efficiency in human evolution were primarily channeled toward further ramping up foraging intensity rather than reducing the energetic costs of subsistence. This unintuitive view of energy use in relation to efficiency finds a parallel in the Jevons Paradox, a macroeconomic principle by which the introduction of more efficient technologies leads to increased consumption rather than savings in human systems (*98*).

Our results also provide deeper evolutionary context for understanding modern trends in human time and energy budgets. Exosomatic energy accounts for a relatively minor portion of the "social metabolism" of small-scale societies (*20*). For example, hunter-gatherers and horticulturalists rely directly or indirectly on biomass generated by solar energy and, with the notable exception of occasional landscape burning practiced in some cultures (*99*), do not participate in systematic large-scale management of ecosystems (*100*). The intensification of agriculture introduced greater exosomatic inputs, primarily in the form of domesticated animals for draft power. Since the Industrial Revolution, fossil fuels and mechanization have increasingly externalized energy production (*101*). Paralleling these changes, the ratio of exosomatic to endosomatic energy flows has risen from less than 5 in hunter-gatherer societies to more than 90 in highly developed industrialized societies (*102*). This has allowed for an unprecedented increase in the energy return on investment of labor (ratio of food energy produced to endosomatic energy invested in labor) for modern agriculture since the 1950s (*103*–*106*). With the subsequent decoupling of industrial production from human and animal labor, industrialized populations have continued to experience reductions in the time costs of "subsistence." For example, the proportion of income spent on food for Americans decreased from ~25% to 12% between 1928 and 1998 (~1.4 and 0.7 hours/day, respectively, assuming a 40-hour work week), mainly due to lowered monetary costs of food (*107*). With large increases in food production alongside increasingly sedentary lifestyles, humans have experienced a fundamental shift in our relationship with energy, setting up one of the major health challenges of our time: the rise of chronic noncommunicable "diseases of civilization" such as obesity, metabolic syndrome, and cardiovascular disease. Unburdened by the high physiological costs of food production, a human body that evolved to expend large quantities of energy to acquire food has now found itself in a potentially deadly mismatch.

Materials and methods

Foraging energetics of nonhuman great apes

To calculate energy budgets of nonhuman great apes, total daily energy expenditure (TEE) was used as a proxy for daily energy acquired from food (kcal/day) under the realistic assumption that energy input and output are approximately equal among nonprovisioned animals in energy balance (*2*, *27*). Food sharing and provisioning are very rare among nonhuman apes in the wild, even between mothers and offspring (*2*, *29*–*31*), and therefore each individual's average daily food energy acquisition must match their average TEE. A lack of surplus production in other great apes is underscored by the fact that humans exhibit elevated fat deposition compared to chimpanzees and gorillas, and that the fat reserves of orangutans fluctuate in accordance with boom and bust seasonal cycles and supra-annual mast fruiting events (*1*, *108*, *109*). TEE was determined for each great ape species by fitting regressions to empirically measured TEE and body mass data for healthy adults (10+ years old) from DLW studies in zoo and sanctuary ape populations (*1*, *110*). We note that TEE for captive primate populations does not differ from that of wild populations in analyses accounting for body size (*27*, *111*). Natural logarithm–transformed values were used for mass and TEE because previous work has demonstrated that, as in other species, TEE increases in a power-law manner with body size in apes (*1*). Regressions were as follows (data presented in fig. S6):

Chimpanzees: $\ln(TEE) = 0.602 \pm 0.196 \ln(mass) + 5.197 \pm 0.792$ (model: adj. $r^2 = 0.23$, $p = 0.005$, SE = 0.195, df = 28)

Gorillas: $\ln(TEE) = 0.726 \pm 0.160 \ln(mass) + 4.432 \pm 0.741$ (model: adj. $r^2 = 0.66$, $p = 0.001$, SE = 0.212, df = 9)

Orangutans: $\ln(TEE) = 0.467 \pm 0.128 \ln(mass) + 5.402 \pm 0.544$ (model: adj. $r^2 = 0.34$, $p = 0.001$, SE = 0.223, df = 23)

We used these regressions to calculate mean (95% CI) TEE for wild males and females of each species using adult body masses reported for wild populations [chimpanzees: males 40.4, females 32.8 (*112*); gorillas (Western lowland): males 170.4, females 70.5 (*113*); orangutans (Borneo): males 78.5, females 35.8 (*113*)]. These estimates for TEE were, in turn, used as estimates of daily energy acquisition, assuming that food provisioning and storage among adult nonhuman great apes is negligible.

To calculate energy expenditure associated with foraging in each species, measurements of average daily distances of terrestrial travel and arboreal climbing were compiled from prior studies (*80*, *114*–*117*) and distances were converted to energy costs (kcal/day) using

established cost of transport values (kcal kg^{-1} m^{-1}) from respirometry studies of chimpanzee quadrupedal walking and nonhuman primate climbing (*16*, *33*). The kinematics and locomotor anatomy of nonhuman great apes are broadly similar across species (*118*), and thus walking costs (kcal kg^{-1} m^{-1}) determined from respirometry studies of chimpanzees also provide a reasonable estimate for those of gorillas and orangutans. Further, the cost of transport (kcal kg^{-1} m^{-1}) for climbing is similar across primates, and indeed across mammals (*33*), and thus the cost of transport for climbing derived from respirometry studies in nonhuman primates provides a reliable estimate of climbing costs in great apes. The time cost of foraging was defined as the daily time spent feeding and moving to acquire food and was calculated by sex for each species by averaging data from prior studies (table S5). Estimates of TEE and behavioral data on foraging in nonhuman great apes were derived from different samples.

To check our regression-based approach for estimating TEE, we compared our estimates to those from (i) activity budget analyses for wild great apes and other primates (*26*), (ii) DLW measurements of wild primates (*27*), (iii) DLW measurements of wild nonprimate mammals (*28*), and (iv) estimated daily energy acquisition values from studies observing feeding behavior in wild great apes (table S1) (*119*–*128*). To facilitate comparisons across the wide range of body sizes in these analyses, we examined the ratio of TEE to basal metabolic rate (BMR; kcal/day), which provides both a body size–adjusted measure for comparison of TEE and a rough measure of daily physical activity (the TEE/BMR ratio is sometimes termed "physical activity level") (*28*, *129*). BMRs were estimated from body mass using published regression equations for nonhuman great apes and other primates (*1*, *26*) (table S1). Our regression-based TEE estimates yielded TEE/BMR values for wild great apes that were similar to activity budget–based estimates, DLW measurements in wild primates, and DLW measurements in wild nonprimate mammals (table S1 and fig. S1). The agreement between our approach and other measures and estimates of daily expenditure supports our use of DLW-based regression estimates of TEE here. However, estimates of food acquisition from behavioral observations of wild great apes (*119*, *124*, *126*, *127*), in at least some studies, yield higher estimates of TEE and TEE/BMR (table S1 and fig. S1). We note that estimates of energy acquisition from feeding observations require estimates of intake rate for each food, average nutritional content of each food, digestibility of each food, and average daily feeding time across partial-day follows, and that analytical error in each of these estimates can lead to compounded errors in estimated TEE; we therefore favor the DLW regression-based approach used in this study. Nonetheless, as a sensitivity analysis, we reran analyses using energy acquisition estimates from wild nonhuman great ape feeding studies. Even if energy returns were somewhat greater than our estimates, as suggested by estimated return rates from feeding observations in wild nonhuman great apes, the pattern of results is unchanged: human foraging efficiencies (F) would compare even more unfavorably to those of other great apes, while hourly return rates remain higher for humans (figs. S3 and S4).

Measurements of subsistence energetics among Hadza hunter-gatherers

To calculate energy expenditure during subsistence-related activities among the Hadza, a portable respirometry system (Cosmed, K4b2) was used to measure breath-by-breath oxygen consumption and carbon dioxide production. Hadza participants (total: $n_{male} = 9$, $n_{female} = 5$) performed various subsistence tasks for 5 to 10 min while wearing the respirometer, and average rates of energy expenditure for each task were calculated once steady-state energy expenditure had been reached. Mass-specific energy expenditure (J kg^{-1} m^{-1}) was converted to net energy cost (kcal/min) by multiplying by the caloric coefficient (20.1) and body mass (kg), and then subtracting the participant-specific resting metabolic rate (i.e., the energy cost of rest in a sitting position). Resting energy expenditure was measured for 5 to 8 min immediately prior to the work task measurement with the subject either standing (for climbing, chopping, pounding, and walking) or kneeling (for digging). For activities lacking respirometry measurements (e.g., tool manufacture, eating) we used values from the literature for similar tasks (table S6).

To assess time spent on subsistence-related activities, scan samples of Hadza adults [n_{male} = 135 (26,498 observations), n_{female} = 179 (37,433 observations)] were collected in 16 camps between 1995–1996 and 2003–2005. Scan-sampling data were collected across all seasons between 7 a.m. and 7 p.m. To generate estimates of time spent in different subsistence activities, we used Bayesian multilevel, multinomial logistic regression models on time allocation data (*130*). In short, this technique models the probability of an individual engaging in a specific behavior (the multinomial response outcome) as a function of independent variables while accounting for repeated observations of individuals and correlated random effects that characterize individual-level trade-offs in the probability of engaging in different behaviors. In our analysis, outcome behaviors were divided into categories representing the major subsistence activities (with different per–unit time costs), and we included age, age^2, time of day, and time of day^2 as fixed effects, as well as random intercepts and their correlations for individual, community, and month. Men and women were analyzed separately. Models were fit using Hamiltonian Monte Carlo algorithms in the *RStan* package in R 3.5.0. To improve mixing of the Hamiltonian Monte Carlo chains, fixed effects were centered and scaled prior to analysis, and weakly informative priors for the fixed effects parameters were used.

After ensuring proper mixing and convergence of the models, model estimates were converted to sex- and age-specific probabilities of engaging in an activity using a custom link script and the softmax function (which normalizes the K predicted probabilities to sum to 1), with random effects set to 0. Given that sampling took place between 7 a.m. and 7 p.m. (when the vast majority of subsistence activity occurs), probability functions for each activity were integrated across time of day to calculate the number of minutes spent on a given activity per day by age and sex.

Because scan sampling occurred within camps, a large proportion of observations fell into a generic activity category for work outside of camp. To categorize out-of-camp activities, we used a database of focal follows in which observers followed individuals on out-of-camp trips and continuously recorded behavior (figs. S7 and S8). Men's time allocation out of camp was based on 46 focal follows collected between 2006 and 2014 among 27 different men (mean age = 33, SD = 11) in eight residential camps. On average, individual focal follows for men lasted 5.3 hours (SD = 2.8). The out-of-camp follow data were categorized into times spent walking, running, chopping, digging, resting, and in other activities (fig. S7 and table S7). "Other activities" include generally low- to moderate-level physical activities, including lying in wait while hunting, scanning the landscape for animals, inspecting trees for bee nests, processing foods, and eating. The energetic cost of "other activities" was ascribed an average value for non-baobab food processing [1.9 kcal/min, for average category of "food preparation" in (*131*)]. We also used observations from these follows to estimate the mean height climbed per day to extract honey (10 m/day).

Hadza women's out-of-camp time allocation while foraging was recorded during 27 focal follows of 14 women (mean age = 45, SD = 14) collected between 2011 and 2014 in three residential camps. Follows of women foraging lasted on average 3.7 hours out of camp (SD = 2). Time allocation during these follows has been categorized into times spent walking, running, chopping, digging, resting, and in other activities (fig. S8 and table S4). "Other activities" are generally low- to moderate-level physical activities and were also ascribed the energetic cost of non-baobab food processing (as above).

During focal follows of both men and women, we recorded every time the observed Hadza subject consumed food, noting the species of food, its directly weighed or visually estimated amount (e.g., volume of honey, numbers of berries, mass of meat), and the source of the food. Analyzing these data, we find that men and women consumed, on average, an additional 306 kcal/hour (SD = 479) and 70 kcal/hour (SD = 60), respectively, while out of camp.

To estimate task-specific daily energy expenditure, respirometry measurements for specific activities were combined with daily time allocation data from scan sampling and focal follows. Total time spent foraging was then calculated by summing time spent on subsistence behaviors both in and out of camp.

Food acquisition in human populations cannot be adequately assessed using TEE from DLW measurements because frequent provisioning and sharing occur within the context of surplus production by some individuals. Food acquisition for the Hadza was estimated using data on foraging in which daily activities were recorded and all food brought to camp was either directly weighed or visually estimated. Data were collected between 2005 and 2009 on a total of 100 women and 79 men from seven camps, and were integrated with caloric density values for each food to calculate the total daily caloric value of food acquisition. Caloric density values of Hadza foods were estimated based on published nutritional values for honey, berries, tubers, baobab, and marula nuts (*68*, *132*, *133*). For Hadza foods where no published values were available, we used USDA values for closely related foods (*134*). These included quail eggs as a proxy for crested francolin eggs and beet greens as a proxy for leafy greens. To estimate the caloric value of animal carcasses, we used an intermediate value of 1.7 kcal/g for all species.

To generate production curves for the Hadza, foraging returns (kcal/day) were modeled as a function of age (third-order polynomial), sex, and age × sex interaction, with random intercepts for individual, camp, and month. Given the large number of zeros and highly skewed distribution of foraging returns, we used a Bayesian lognormal hurdle model (a commonly used alternative, the gamma hurdle model, produced similar results) (fig. S5) (*135*). Estimation was performed using the *brms* package in R using weakly informative priors (run with four chains with 3000 iterations, 1500 of which were devoted to warmup). Finally, total daily acquisition was obtained by summing in-camp returns with out-of-camp consumption (by multiplying the number of hours spent out of camp and the average per-hour consumption estimates above, scaled to the productivity age profile).

TEE among the Hadza was measured in a sample of 27 adult men and 24 adult women using the doubly labeled water method (*136*). Fat-free mass was measured by multiplying average body mass by 1 – (% body fat), with percent body fat measured using bioimpedance.

Measurements of subsistence energetics among Tsimane horticulturalists

Energy expenditure during subsistence-related activities among the Tsimane was measured using the same respirometry procedures used among the Hadza but applied to Tsimane subsistence tasks (Tsimane participants: n_{male} = 7, n_{female} = 5). Resting energy expenditures for Tsimane were similarly measured in standing or sitting position prior to performing subsistence activities.

To estimate time spent in subsistence activities, we analyzed adult time allocation data (n = 35,500 observations) collected from 2002 to 2007 as part of a longitudinal study of the Tsimane (*137*). Time allocation data represent instantaneous scan samples collected at 30-min intervals. Villages were partitioned into household clusters that were monitored for 2- or 3-hour time blocks between 7 a.m. and 7 p.m. If a residing individual was not present at the time of sampling, interviews with other residents were used to assess out-of-camp activities (follow-up direct observation indicated that reported behaviors were highly accurate). Data were collected across all seasons in nine villages. Analyses were restricted to adults between the ages of 15 and 75 (n_{male} = 282, n_{female} = 243), and the same analysis and post-processing procedures reported for Hadza above were employed using multilevel, multinomial logistic regression models to calculate time spent in different activities. Finally, time spent in each subsistence activity was multiplied by net energy costs from respirometry to estimate the daily cost (kcal) of subsistence (E_f).

For some subsistence tasks (e.g., rice processing by men), there were not enough observations to accurately estimate separate parameters in the models. In such cases, outcomes were combined into the most closely related category and energetic costs of those activities were recalculated as a weighted mean based on the raw proportion of each observed activity (e.g., if rice processing has a net cost of 300 kcal/hour, general processing costs 100 kcal/hour, and rice processing accounts for one-fourth of observations in either category, then the combined cost assigned would be 150 kcal/hour).

Estimates of Tsimane gross energy production (E_a, kcal/day) from foraging and horticulture, separated by sex and age, come from reanalysis of data from previous studies (*138*). TEE among the Tsimane was measured in a sample of 18 adult men and 22 adult women using the doubly labeled water method (*97*). Fat-free mass was measured in the same manner as described for the Hadza.

Global data on human subsistence energetics

To generalize beyond the Hadza and Tsimane and to facilitate cross-cultural comparisons, a literature search was conducted for published estimates of food/energy acquisition from subsistence activities, the energetic costs of subsistence activities, time devoted to subsistence, foraging efficiency, and return rates in hunter-gatherer and horticulturalist societies around the world (table S2). Values were disaggregated by sex whenever possible, but in many instances were available only in combination (e.g., many horticulturalist production systems, where it is difficult to disentangle production within a household unit). In order to be included in our analysis, a study needed to present at least one of the quantities of interest (E_a, E_f, T_f, F, $R_{g/n}$) in a form that was commensurate to those estimated for our study populations. Return rates were calculated to include processing whenever possible. Time allocation estimates for other auxiliary subsistence activities, however, including tool manufacturing, eating, and water and firewood collection, were unavailable for most societies besides the Hadza and Tsimane and thus were not included in the cross-cultural analyses. A detailed description of how all values in the cross-cultural sample were derived is provided in the additional methods.

REFERENCES AND NOTES

1. H. Pontzer *et al.*, Metabolic acceleration and the evolution of human brain size and life history. *Nature* **533**, 390–392 (2016). doi: 10.1038/nature17654; pmid: 27144364
2. H. Kaplan, K. Hill, J. Lancaster, A. M. Hurtado, A theory of human life history evolution: Diet, intelligence, and longevity. *Evol. Anthropol.* **9**, 156–185 (2000). doi: 10.1002/1520-6505(2000)9:4<156::AID-EVAN5>3.0.CO;2-7
3. D. E. Lieberman, *The Story of the Human Body: Evolution, Health, and Disease* (Pantheon, 2013).
4. J. P. Bocquet-Appel, S. Naji, Testing the hypothesis of a worldwide Neolithic demographic transition: Corroboration from American cemeteries. *Curr. Anthropol.* **47**, 341–365 (2006). doi: 10.1086/498948
5. R. C. Ydenberg, C. V. J. Welham, R. Schmid-Hempel, P. Schmid-Hempel, G. Beauchamp, Time and energy constraints and the relationships between currencies in foraging theory. *Behav. Ecol.* **5**, 28–34 (1994). doi: 10.1093/beheco/5.1.28
6. A. Kacelnik, A. I. Houston, Some effects of energy costs on foraging strategies. *Anim. Behav.* **32**, 609–614 (1984). doi: 10.1016/S0003-3472(84)80298-5
7. P. A. Abrams, Optimal traits when there are several costs: The interaction of mortality and energy costs in determining foraging behavior. *Behav. Ecol.* **4**, 246–259 (1993). doi: 10.1093/beheco/4.3.246
8. E. A. Smith, *Inujjuamiut Foraging Strategies: Evolutionary Ecology of an Arctic Hunting Economy* (Aldine, 1991).
9. D. W. Stephens, J. R. Krebs, *Foraging Theory* (Princeton Univ. Press, 1986).
10. E. A. Smith, Human adaptation and energetic efficiency. *Hum. Ecol. Interdiscip. J.* **7**, 53–74 (1979). doi: 10.1007/BF00889352
11. R. C. Ydenberg, in *Foraging: Behaviour and Ecology* (Univ. of Chicago Press, 2007), pp. 273–303.
12. J. M. McNamara, A. I. Houston, Currencies for foraging based on energetic gain. *Am. Nat.* **150**, 603–617 (1997). doi: 10.1086/286084; pmid: 18811303
13. P. Schmid-Hempel, A. Kacelnik, A. I. Houston, Honeybees maximize efficiency by not filling their crop. *Behav. Ecol. Sociobiol.* **17**, 61–66 (1985). doi: 10.1007/BF00299430

14. A. Kacelnik, A. I. Houston, P. Schmid-Hempel, Central-place foraging in honey bees: The effect of travel time and nectar flow on crop filling. *Behav. Ecol. Sociobiol.* **19**, 19–24 (1986). doi: 10.1007/BF00303838
15. P. Schmid-Hempel, Efficient nectar-collecting by honeybees I. Economic models. *J. Anim. Ecol.* **56**, 209 (1987). doi: 10.2307/4810
16. H. Pontzer, D. A. Raichlen, M. D. Sockol, The metabolic cost of walking in humans, chimpanzees, and early hominins. *J. Hum. Evol.* **56**, 43–54 (2009). doi: 10.1016/j.jhevol.2008.09.001; pmid: 18986682
17. L. Aiello, P. Wheeler, The expensive-tissue hypothesis: The brain and the digestive system in human and primate evolution. *Curr. Anthropol.* **36**, 199–221 (1995). doi: 10.1086/204350
18. C. Organ, C. L. Nunn, Z. Machanda, R. W. Wrangham, Phylogenetic rate shifts in feeding time during the evolution of *Homo*. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 14555–14559 (2011). doi: 10.1073/pnas.1107806108; pmid: 21873223
19. A. J. Lotka, *Elements of Mathematical Biology* (Dover, 1956).
20. M. Fischer-Kowalski, F. Krausmann, I. Pallua, A sociometabolic reading of the Anthropocene: Modes of subsistence, population size and human impact on Earth. *Anthr. Rev.* **1**, 8–33 (2014). doi: 10.1177/2053019613518033
21. F. W. Marlowe, *The Hadza: Hunter-gatherers of Tanzania* (Univ. of California Press, 2010).
22. A. N. Crittenden, S. L. Schnorr, Current views on hunter-gatherer nutrition and the evolution of the human diet. *Am. J. Phys. Anthropol.* **162** (suppl. 63), 84–109 (2017). doi: 10.1002/ajpa.23148; pmid: 28105723
23. A. Benito-Calvo *et al*., 360° surface morphometric analysis of pounding stone tools used by Hadza foragers of Tanzania: A new methodological approach for studying percussive stone artefacts. *J. Archaeol. Sci. Rep.* **20**, 611–621 (2018). doi: 10.1016/j.jasrep.2018.06.003
24. M. Kleiber, *The Fire of Life: An Introduction to Animal Energetics* (Wiley, 1961).
25. J. R. Speakman, *Doubly Labelled Water: Theory & Practice* (Chapman & Hall, 1997).
26. W. R. Leonard, M. L. Robertson, Comparative primate energetics and hominid evolution. *Am. J. Phys. Anthropol.* **102**, 265–281 (1997). doi: 10.1002/(SICI)1096-8644(199702)102:2<265::AID-AJPA8>3.0.CO;2-X; pmid: 9066904
27. H. Pontzer *et al*., Primate energy expenditure and life history. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 1433–1437 (2014). doi: 10.1073/pnas.1316940111; pmid: 24474770
28. K. R. Westerterp, J. R. Speakman, Physical activity energy expenditure has not declined since the 1980s and matches energy expenditures of wild mammals. *Int. J. Obes.* **32**, 1256–1263 (2008). doi: 10.1038/ijo.2008.74; pmid: 18504442
29. R. M. Wittig *et al*., Food sharing is linked to urinary oxytocin levels and bonding in related and unrelated wild chimpanzees. *Proc. R. Soc. B* **281**, 20133096 (2014). doi: 10.1098/rspb.2013.3096; pmid: 24430853
30. A. V. Jaeggi, C. P. Van Schaik, The evolution of food sharing in primates. *Behav. Ecol. Sociobiol.* **65**, 2125–2140 (2011). doi: 10.1007/s00265-011-1221-3
31. A. V. Jaeggi, M. A. van Noordwijk, C. P. van Schaik, Begging for information: Mother-offspring food sharing among wild Bornean orangutans. *Am. J. Primatol.* **70**, 533–541 (2008). doi: 10.1002/ajp.20525; pmid: 18186082
32. H. Pontzer, D. A. Raichlen, P. S. Rodman, Bipedal and quadrupedal locomotion in chimpanzees. *J. Hum. Evol.* **66**, 64–82 (2014). doi: 10.1016/j.jhevol.2013.10.002; pmid: 24315239
33. J. B. Hanna, D. Schmitt, T. M. Griffin, The energetic cost of climbing in primates. *Science* **320**, 898 (2008). doi: 10.1126/science.1155504; pmid: 18487185
34. J. P. Bocquet-Appel, When the world's population took off: The springboard of the Neolithic Demographic Transition. *Science* **333**, 560–561 (2011). doi: 10.1126/science.1208880; pmid: 21798934
35. D. W. Sellen, R. Mace, Fertility and mode of subsistence: A phylogenetic analysis. *Curr. Anthropol.* **38**, 878–889 (1997). doi: 10.1086/204677
36. S. Bowles, Cultivation of cereals by the first farmers was not more productive than foraging. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 4760–4765 (2011). doi: 10.1073/pnas.1010733108; pmid: 21383181
37. D. J. Kennett, B. Winterhalder, *Behavioral Ecology and the Transition to Agriculture* (Univ. of California Press, 2006).
38. K. L. Kramer, P. T. Ellison, Pooled energy budgets: Resituating human energy-allocation trade-offs. *Evol. Anthropol.* **19**, 136–147 (2010). doi: 10.1002/evan.20265
39. M. J. T. Norman, Energy inputs and outputs of subsistence cropping systems in the tropics. *Agro-ecosyst.* **4**, 355–366 (1978). doi: 10.1016/0304-3746(78)90001-X
40. H. Pontzer, Relating ranging ecology, limb length, and locomotor economy in terrestrial animals. *J. Theor. Biol.* **296**, 6–12 (2012). doi: 10.1016/j.jtbi.2011.11.018; pmid: 22138093
41. K. D. Reinhardt, Wild primate sleep: Understanding sleep in an ecological context. *Curr. Opin. Physiol.* **15**, 238–244 (2020). doi: 10.1016/j.cophys.2020.04.002
42. N. F. Butte, J. M. Hopkinson, N. Mehta, J. K. Moon, E. O. Smith, Adjustments in energy expenditure and substrate utilization during late pregnancy and lactation. *Am. J. Clin. Nutr.* **69**, 299–307 (1999). doi: 10.1093/ajcn/69.2.299; pmid: 9989696
43. F. W. Marlowe, A critical period for provisioning by Hadza men: Implications for pair bonding. *Evol. Hum. Behav.* **24**, 217–229 (2003). doi: 10.1016/S1090-5138(03)00014-X
44. M. Gurven, H. Kaplan, A. Zelada Supa, Mortality experience of Tsimane Amerindians of Bolivia: Regional variation and temporal trends. *Am. J. Hum. Biol.* **19**, 376–398 (2007). doi: 10.1002/ajhb.20600; pmid: 17421012
45. N. G. Blurton Jones, *Demography and Evolutionary Ecology of Hadza Hunter-Gatherers* (Cambridge Univ. Press, 2016).
46. K. L. Kramer, The cooperative economy of food: Implications for human life history and physiology. *Physiol. Behav.* **193** (Pt B), 196–204 (2018). doi: 10.1016/j.physbeh.2018.03.029; pmid: 29630962
47. C. W. Kuzawa *et al*., Metabolic costs and evolutionary implications of human brain development. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 13010–13015 (2014). doi: 10.1073/pnas.1323099111; pmid: 25157149
48. R. Hames, in *Risk and Uncertainty in Tribal and Peasant Economies*, E. Cashdan, Ed. (Westview, 1990), pp. 89–106.
49. A. M. Hurtado, K. Hill, Early dry season subsistence ecology of the Cuiva (Hiwi) foragers of Venezuela. *Hum. Ecol. Interdiscip. J.* **15**, 163–187 (1987). doi: 10.1007/BF00888379
50. K. Hill, H. Kaplan, K. Hawkes, M. Hurtado, Foraging decisions among Aché hunter-gatherers: New data and implications for optimal foraging models. *Ethol. Sociobiol.* **8**, 1–36 (1987). doi: 10.1016/0162-3095(87)90055-0
51. W. F. Keegan, The optimal foraging analysis of horticultural production. *Am. Anthropol.* **88**, 92–107 (1986). doi: 10.1525/aa.1986.88.1.02a00060
52. D. R. Piperno, in *Behavioral Ecology and the Transition to Agriculture*, D. J. Kennett, B. Winterhalder, Eds. (Univ. of California Press, 2006), pp. 137–166.
53. R. Hames, A comparison of the efficiencies of the shotgun and bow in neotropical forest hunting. *Hum. Ecol. Interdiscip. J.* **7**, 219–252 (1979). doi: 10.1007/BF00889493
54. D. S. Wilkie, B. Curran, Why do Mbuti hunters use nets? Ungulate hunting efficiency of archers and net-hunters in the Ituri Rain Forest. *Am. Anthropol.* **93**, 680–689 (1991). doi: 10.1525/aa.1991.93.3.02a00080
55. A. M. Hurtado, K. Hill, Experimental studies of tool efficiency among Machiguenga women and implications for root-digging foragers. *J. Anthropol. Res.* **45**, 207–217 (1989). doi: 10.1086/jar.45.2.3630334
56. A. Ugan, J. Bright, A. Rogers, When is technology worth the trouble? *J. Archaeol. Sci.* **30**, 1315–1329 (2003). doi: 10.1016/S0305-4403(03)00022-0
57. N. E. Stevens, R. McElreath, When are two tools better than one? Mortars, millingslabs, and the California acorn economy. *J. Anthropol. Archaeol.* **37**, 100–111 (2015). doi: 10.1016/j.jaa.2014.12.002
58. J. A. Yost, P. M. Kelley, in *Adaptive Responses of Native Amazonians*, R. Hames, W. Vickers, Eds. (Academic Press, 1983), pp. 189–224.
59. R. Salisbury, *From Stone to Steel* (Cambridge Univ. Press, 1962).
60. W. R. Leonard, M. L. Robertson, Nutritional requirements and human evolution: A bioenergetics model. *Am. J. Hum. Biol.* **4**, 179–195 (1992). doi: 10.1002/ajhb.1310040204; pmid: 28524347
61. D. E. Lieberman, *Exercised: Why Something We Never Evolved to Do Is Healthy and Rewarding* (Pantheon, 2020).
62. K. Isler, C. van Schaik, Costs of encephalization: The energy trade-off hypothesis tested on birds. *J. Hum. Evol.* **51**, 228–243 (2006). doi: 10.1016/j.jhevol.2006.03.006; pmid: 16730368
63. A. Navarrete, C. P. van Schaik, K. Isler, Energetics and the evolution of human brain size. *Nature* **480**, 91–93 (2011). doi: 10.1038/nature10629; pmid: 22080949
64. A. I. Houston, Energetic constraints and foraging efficiency. *Behav. Ecol.* **6**, 393–396 (1995). doi: 10.1093/beheco/6.4.393
65. B. Simmen *et al*., Primate energy input and the evolutionary transition to energy-dense diets in humans. *Proc. R. Soc. B* **284**, 20170577 (2017). doi: 10.1098/rspb.2017.0577; pmid: 28592672
66. K. D. Zink, D. E. Lieberman, Impact of meat and Lower Palaeolithic food processing techniques on chewing in humans. *Nature* **531**, 500–503 (2016). doi: 10.1038/nature16990; pmid: 26958832
67. R. N. Carmody, G. S. Weintraub, R. W. Wrangham, Energetic consequences of thermal and nonthermal food processing. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 19199–19203 (2011). doi: 10.1073/pnas.1112128108; pmid: 22065771
68. J. C. Berbesque, B. M. Wood, A. N. Crittenden, A. Mabulla, F. W. Marlowe, Eat first, share later: Hadza hunter-gatherer men consume more while foraging than in central places. *Evol. Hum. Behav.* **37**, 281–286 (2016). doi: 10.1016/j.evolhumbehav.2016.01.003
69. G. F. Oster, E. O. Wilson, *Caste and Ecology in the Social Insects* (Princeton Univ. Press, 1978).
70. D. W. Davidson, The role of resource imbalances in the evolutionary ecology of tropical arboreal ants. *Biol. J. Linn. Soc. London* **61**, 153–181 (1997). doi: 10.1111/j.1095-8312.1997.tb01785.x
71. U. G. Mueller, S. A. Rehner, T. R. Schultz, The evolution of agriculture in ants. *Science* **281**, 2034–2038 (1998). doi: 10.1126/science.281.5385.2034; pmid: 9748164
72. F. Dyer, T. Seeley, Nesting behavior and the evolution of worker tempo in four honey bee species. *Ecology* **72**, 156–170 (1991). doi: 10.2307/1938911
73. C. G. Farmer, Parental care: The key to understanding endothermy and other convergent features in birds and mammals. *Am. Nat.* **155**, 326–334 (2000). doi: 10.1086/303323; pmid: 10718729
74. S. B. Hrdy, J. M. Burkart, The emergence of emotionally modern humans: Implications for language and learning. *Philos. Trans. R. Soc. London Ser. B* **375**, 20190499 (2020). doi: 10.1098/rstb.2019.0499; pmid: 32475330
75. S. B. Hrdy, *Mothers and Others: The Evolutionary Origins of Mutual Understanding* (Harvard Univ. Press, 2009).
76. J. M. Burkart, S. B. Hrdy, C. P. Van Schaik, Cooperative breeding and human cognitive evolution. *Evol. Anthropol.* **18**, 175–186 (2009). doi: 10.1002/evan.20222
77. K. Hawkes, J. F. O'Connell, N. G. Blurton Jones, O. T. Oftedal, R. J. Blumenschine, Hunting income patterns among the Hadza: Big game, common goods, foraging goals and the evolution of the human diet. *Philos. Trans. R. Soc. London Ser. B* **334**, 243–250 (1991). doi: 10.1098/rstb.1991.0113; pmid: 1685582
78. B. M. Wood *et al*., Gendered movement ecology and landscape use in Hadza hunter-gatherers. *Nat. Hum. Behav.* **5**, 436–446 (2021). doi: 10.1038/s41562-020-01002-7; pmid: 33398143
79. K. Isler, C. P. van Schaik, The Expensive Brain: A framework for explaining evolutionary changes in brain size. *J. Hum. Evol.* **57**, 392–400 (2009). doi: 10.1016/j.jhevol.2009.04.009; pmid: 19732937
80. H. Pontzer, R. W. Wrangham, Climbing and the daily energy cost of locomotion in wild chimpanzees: Implications for hominoid locomotor evolution. *J. Hum. Evol.* **46**, 317–335 (2004). doi: 10.1016/j.jhevol.2003.12.006; pmid: 14984786
81. H. Pontzer, Economy and endurance in human evolution. *Curr. Biol.* **27**, R613–R621 (2017). doi: 10.1016/j.cub.2017.05.031; pmid: 28633035
82. D. R. Carrier, The energetic paradox of human running and hominid evolution. *Curr. Anthropol.* **25**, 483–495 (1984). doi: 10.1086/203165
83. R. E. Shave *et al*., Selection of endurance capabilities and the trade-off between pressure and volume in the evolution of the human heart. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 19905–19910 (2019). doi: 10.1073/pnas.1906902116; pmid: 31527253
84. R. N. Carmody, R. W. Wrangham, The energetic significance of cooking. *J. Hum. Evol.* **57**, 379–391 (2009). doi: 10.1016/j.jhevol.2009.02.011; pmid: 19732938
85. S. L. Schnorr, A. N. Crittenden, K. Venema, F. W. Marlowe, A. G. Henry, Assessing digestibility of Hadza tubers using a dynamic in-vitro model. *Am. J. Phys. Anthropol.* **158**, 371–385 (2015). doi: 10.1002/ajpa.22805; pmid: 26174414
86. A. Stahl, in *Foraging and Farming: The Evolution of Plant Exploitation*, D. Harris, G. Hillman, Eds. (Unwin Hyman, 1989), pp. 171–194.

87. S. M. Boback *et al*., Cooking and grinding reduces the cost of meat digestion. *Comp. Biochem. Physiol. A* **148**, 651–656 (2007). doi: 10.1016/j.cbpa.2007.08.014; pmid: 17827047
88. R. Hames, in *Evolutionary Ecology and Human Behavior*, E. A. Smith, B. Winterhalder, Eds. (de Gruyter, 1992), pp. 203–237.
89. P. Shipman, A. Walker, The costs of becoming a predator. *J. Hum. Evol.* **18**, 373–392 (1989). doi: 10.1016/0047-2484(89)90037-7
90. J. Henrich, *The Secret of Our Success: How Culture is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter* (Princeton Univ. Press, 2016).
91. K. R. Hill, B. M. Wood, J. Baggio, A. M. Hurtado, R. T. Boyd, Hunter-gatherer inter-band interaction rates: Implications for cumulative culture. *PLOS ONE* **9**, e102806 (2014). doi: 10.1371/journal.pone.0102806; pmid: 25047714
92. J. T. Faith, J. Rowan, A. Du, P. L. Koch, Plio-Pleistocene decline of African megaherbivores: No evidence for ancient hominin impacts. *Science* **362**, 938–941 (2018). doi: 10.1126/science.aau2728; pmid: 30467167
93. M. D. Sahlins, in *Man the Hunter*, R. B. Lee, I. DeVore, Eds. (Aldine, 1968), pp. 85–89.
94. J. Suzman, *Affluence Without Abundance: What We Can Learn from the World's Most Successful Civilisation* (Bloomsbury, 2019).
95. M. Dyble, J. Thorley, A. E. Page, D. Smith, A. B. Migliano, Engagement in agricultural work is associated with reduced leisure time among Agta hunter-gatherers. *Nat. Hum. Behav.* **3**, 792–796 (2019). doi: 10.1038/s41562-019-0614-6; pmid: 31110340
96. A. E. Page *et al*., Reproductive trade-offs in extant hunter-gatherers suggest adaptive mechanism for the Neolithic expansion. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 4694–4699 (2016). doi: 10.1073/pnas.1524031113; pmid: 27071109
97. M. D. Gurven *et al*., High resting metabolic rate among Amazonian forager-horticulturalists experiencing high pathogen burden. *Am. J. Phys. Anthropol.* **161**, 414–425 (2016). doi: 10.1002/ajpa.23040; pmid: 27375044
98. B. Alcott, Jevons' paradox. *Ecol. Econ.* **54**, 9–21 (2005). doi: 10.1016/j.ecolecon.2005.03.020
99. D. W. Bird, R. B. Bird, B. F. Codding, N. Taylor, A landscape architecture of fire: Cultural emergence and ecological pyrodiversity in Australia's Western Desert. *Curr. Anthropol.* **57**, S65–S79 (2016). doi: 10.1086/685763
100. F. Krausmann, H. Weisz, N. Eisenmenger, in *Social Ecology* (Springer, 2016), pp. 63–92.
101. H. Pontzer, Hotter and sicker: External energy expenditure and the tangled evolutionary roots of anthropogenic climate change and chronic disease. *Am. J. Hum. Biol.* **33**, e23579 (2021). doi: 10.1002/ajhb.23579; pmid: 33629785
102. M. Giampietro, K. Mayumi, Multiple-scale integrated assessment of societal metabolism: Introducing the approach. *Popul. Environ.* **22**, 109–153 (2000). doi: 10.1023/A:1026691623300
103. V. Smil, *Energy in Nature and Society: General Energetics of Complex Systems* (MIT Press, 2008).
104. S. Harchaoui, P. Chatzimpiros, Energy, nitrogen, and farm surplus transitions in agriculture from historical data modeling. France, 1882–2013. *J. Ind. Ecol.* **23**, 412–425 (2019). doi: 10.1111/jiec.12760
105. S. Gingrich, F. Krausmann, At the core of the socio-ecological transition: Agroecosystem energy fluxes in Austria 1830-2010. *Sci. Total Environ.* **645**, 119–129 (2018). doi: 10.1016/j.scitotenv.2018.07.074; pmid: 30016707
106. D. Pimentel *et al*., Food production and the energy crisis. *Science* **182**, 443–449 (1973). doi: 10.1126/science.182.4111.443; pmid: 17832454
107. A. Drewnowski, Fat and sugar: An economic analysis. *J. Nutr.* **133**, 838S–840S (2003). doi: 10.1093/jn/133.3.838S; pmid: 12612164
108. C. D. Knott, in *Seasonality in Primates: Studies of Living and Extinct Human and Non-Human Primates*, D. Brockman, C. van Schaik, Eds. (Cambridge Univ. Press, 2005). 351–378.
109. K. Isler, in *Adipose Tissue and Adipokines in Health and Disease* (Humana, 2014), pp. 3–13.
110. H. Pontzer *et al*., Evolution of water conservation in humans. *Curr. Biol.* **31**, 1804–1810.e5 (2021). doi: 10.1016/j.cub.2021.02.045; pmid: 33675699
111. H. Pontzer, The crown joules: Energetics, ecology, and evolution in humans and other primates. *Evol. Anthropol.* **26**, 12–24 (2017). doi: 10.1002/evan.21513; pmid: 28233387
112. M. L. Carter, H. Pontzer, R. W. Wrangham, J. K. Peterhans, Skeletal pathology in *Pan troglodytes schweinfurthii* in Kibale National Park, Uganda. *Am. J. Phys. Anthropol.* **135**, 389–403 (2008). doi: 10.1002/ajpa.20758; pmid: 18067120
113. R. J. Smith, W. L. Jungers, Body mass in comparative primatology. *J. Hum. Evol.* **32**, 523–559 (1997). doi: 10.1006/jhev.1996.0122; pmid: 9210017
114. E. Kozma, thesis, City University of New York (2020).
115. J. Yamagiwa, A. K. Basabose, in *Primates and Cetaceans*, J. Yamagiwa, L. Karczmarksi, Eds. (Springer, 2014), pp. 43–74.
116. I. Singleton, C. D. Knott, H. C. Morrogh-Bernard, S. A. Wich, C. P. van Schaik, in *Orangutans: Geographic Variation in Behavioral Ecology and Conservation*, S. Wich, S. Utami Atmoko, T. Mitra Setia, C. van Schaik, Eds. (Oxford Univ. Press, 2010), pp. 205–213.
117. H. Pontzer, in *Chimpanzees and Human Evolution*, M. Muller, R. W. Wrangham, D. Pilbeam, Eds. (Harvard Univ. Press, 2017), pp. 259–285.
118. E. M. Finestone, M. H. Brown, S. R. Ross, H. Pontzer, Great ape walking kinematics: Implications for hominoid evolution. *Am. J. Phys. Anthropol.* **166**, 43–55 (2018). doi: 10.1002/ajpa.23397; pmid: 29313896
119. N. L. Conklin-Brittain, C. D. Knott, R. Wrangham, in *Feeding Ecology in Apes and Other Primates*, G. Hohmann, M. Robbins, C. Boesch, Eds. (Cambridge Univ. Press, 2006), pp. 445–571.
120. E. R. Vogel *et al*., Nutritional ecology of wild Bornean orangutans (*Pongo pygmaeus wurmbii*) in a peat swamp habitat: Effects of age, sex, and season. *Am. J. Primatol.* **79**, 1–20 (2017). doi: 10.1002/ajp.22618; pmid: 27889926
121. M. Harrison, H. Morrogh-Bernard, D. J. Chivers, Orangutan energetics and the influence of fruit availability in the nonmasting peat-swamp forest of Sabangau, Indonesian Borneo. *Int. J. Primatol.* **31**, 585–607 (2010). doi: 10.1007/s10764-010-9415-5
122. A. DiGiorgio, thesis, Boston University (2019).
123. J. M. Rothman, E. S. Dierenfeld, H. F. Hintz, A. N. Pell, Nutritional quality of gorilla diets: Consequences of age, sex, and season. *Oecologia* **155**, 111–122 (2008). doi: 10.1007/s00442-007-0901-1; pmid: 17999090
124. S. Masi *et al*., The influence of seasonal frugivory on nutrient and energy intake in wild western gorillas. *PLOS ONE* **10**, e0129254 (2015). doi: 10.1371/journal.pone.0129254; pmid: 26154509
125. J. Lodwick, thesis, Stony Brook University (2014).
126. P. D. Valé *et al*., Energetic management in wild chimpanzees (*Pan troglodytes verus*) in Taï National Park, Côte d'Ivoire. *Behav. Ecol. Sociobiol.* **75**, 1–18 (2021). doi: 10.1007/s00265-020-02935-9
127. M. Uwimbabazi *et al*., Influence of fruit availability on macronutrient and energy intake by female chimpanzees. *Afr. J. Ecol.* **57**, 454–465 (2019). doi: 10.1111/aje.12636; pmid: 32863473
128. S. Lindshield, thesis, Iowa State University (2014).
129. C. Thurber *et al*., Extreme events reveal an alimentary limit on sustained maximal human energy expenditure. *Sci. Adv.* **5**, eaaw0341 (2019). doi: 10.1126/sciadv.aaw0341; pmid: 31183404
130. J. Koster, R. McElreath, Multinomial analysis of behavior: Statistical methods. *Behav. Ecol. Sociobiol.* **71**, 138 (2017). doi: 10.1007/s00265-017-2363-8; pmid: 28959087
131. E. Montgomery, A. Johnson, Machiguenga energy expenditure. *Ecol. Food Nutr.* **6**, 97–105 (1977). doi: 10.1080/03670244.1977.9990487
132. S. S. Murray, M. J. Schoeninger, H. T. Bunn, T. R. Pickering, J. A. Marlett, Nutritional composition of some wild plant foods and honey used by Hadza foragers of Tanzania. *J. Food Compos. Anal.* **14**, 3–13 (2001). doi: 10.1006/jfca.2000.0960
133. M. J. Schoeninger, H. T. Bunn, S. S. Murray, J. A. Marlett, Composition of tubers used by Hadza foragers of Tanzania. *J. Food Compos. Anal.* **14**, 15–25 (2001). doi: 10.1006/jfca.2000.0961
134. USDA, USDA food composition database (2019), https://ndb.nal.usda.gov/.
135. R. McElreath, J. Koster, Using multilevel models to estimate variation in foraging returns. *Hum. Nat.* **25**, 100–120 (2014). doi: 10.1007/s12110-014-9193-4; pmid: 24522975
136. H. Pontzer, B. M. Wood, D. A. Raichlen, Hunter-gatherers as models in public health. *Obes. Rev.* **19** (suppl. 1), 24–35 (2018). doi: 10.1111/obr.12785; pmid: 30511505
137. M. Gurven *et al*., The Tsimane Health and Life History Project: Integrating anthropology and biomedicine. *Evol. Anthropol.* **26**, 54–73 (2017). doi: 10.1002/evan.21515; pmid: 28429567
138. P. L. Hooper, thesis, University of New Mexico (2011).

ACKNOWLEDGMENTS

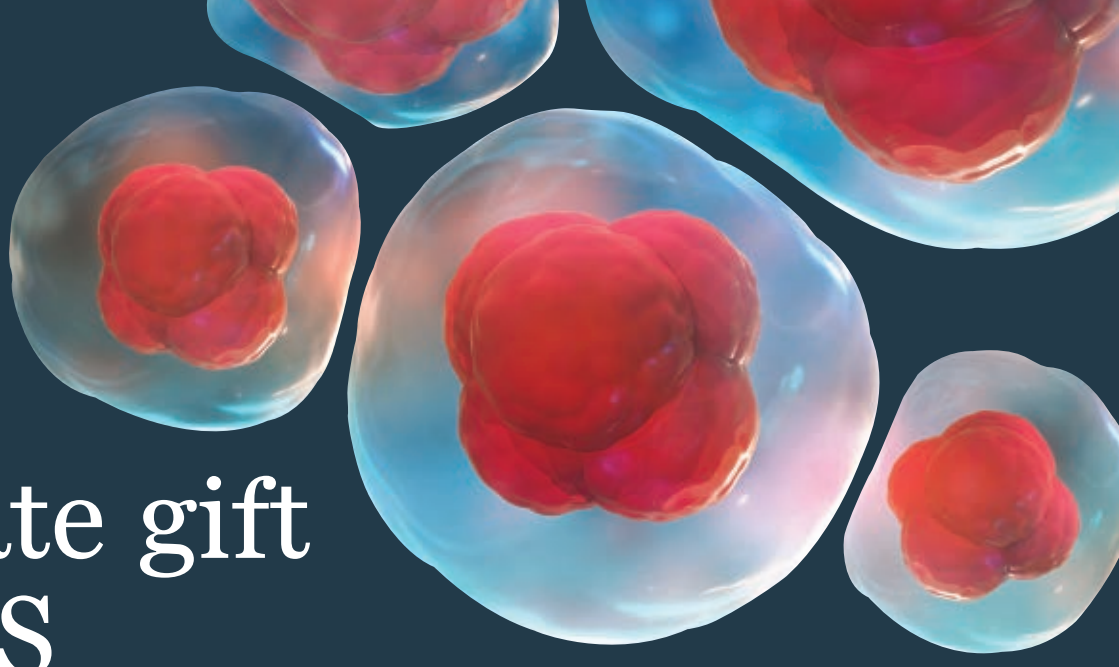
We thank Hadza and Tsimane study participants who generously and patiently allowed us to conduct this research. D. Cummings provided support during fieldwork, J. Koster and R. McElreath provided statistical guidance, and R. Wrangham helped identify data on great ape foraging. **Funding:** Supported by NSF (BCS0422690, BCS-0850815, BCS-1440867, BCS-1062879, BCS-1440841, BCS-1440671, BCS-0242455), NIH (R01AG024119, R56AG024119), the Leakey Foundation, the Max Planck Institute for Evolutionary Anthropology, the University of California, San Diego, and the American School of Prehistoric Research (Harvard University). V.V.V. and J.S. acknowledge IAST funding from ANR under grant ANR-17-EUR-0010 (Investissements d'Avenir program). **Author contributions:** H.P. and T.S.K. conceived the idea. T.S.K. and V.V.V. wrote the paper. T.S.K., V.V.V., I.J.W., N.H., A.N.C., J.S., J.H., D.A.R., B.W., M.G., and H.P. contributed to data collection. T.S.K., V.V.V., B.W., and H.P. analyzed the data. All authors contributed text and ideas and gave final approval for publication. **Competing interests:** The authors of this study declare no competing interests. **Data and materials availability:** Because of existing data-sharing agreements and ethical concerns involving data access, data for Hadza/Tsimane used in this paper are only available from B.W./M.G. under a material transfer agreement with UCLA/UC Santa Barbara. All other data and code (in the R language) for models and analyses are available at https://osf.io/92e6c/. Research with the Tsimane was approved by institutional review boards at the University of New Mexico and UC Santa Barbara, and the Gran Consejo Tsimane. Research with the Hadza was approved by institutional review boards at Harvard University, Yale University, University of California, San Diego, and the University of California, Los Angeles, as well as the Tanzania Commission for Science and Technology (COSTECH). Informed consent was obtained from all participating subjects.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abf0130
Materials and Methods
Tables S1 to S7
Figs. S1 to S8
References (*139*–*204*)
Details of the cross-cultural sample

28 September 2020; accepted 28 October 2021
10.1126/science.abf0130

1848 SOCIETY
AAAS



An estate gift to AAAS

Going all the way back to 1848, our founding year, the American Association for the Advancement of Science (AAAS) has been deeply committed to advancing science, engineering and innovation around the world for the benefit of all people.

Today, we are dedicated to advocating for science and scientific evidence to be fully and positively integrated into public policy and for the community to speak with one voice to advance science and engineering in the United States and around the world.

By making AAAS a beneficiary of your will, trust, retirement plan or life insurance policy, you will become a member of our 1848 Society and will help fuel our work on behalf of science and society – including publishing the world's most promising, innovative research in the *Science* family of journals and engaging in the issues that matter locally, nationally and around the world.

"As a teacher and instructor, I bear responsibility for the younger generations. If you have extra resources, concentrate them on organizations, like AAAS, that are doing work for all."

—Prof. Elisabeth Ervin-Blankenheim, 1848 Society member

If you intend to include AAAS in your estate plans, provide this information to your lawyer or financial adviser:

Legal Name: American Association for the Advancement of Science
Federal Tax ID Number: 53-0196568
Address: 1200 New York Avenue, NW, Washington, DC 20005

If you would like more information on making an estate gift to AAAS, cut out and return the form below or send an email to philanthropy@aaas.org. Additional details are also available online at **www.aaas.org/1848Society**.

AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE

cut here

Yes, I would like more information about joining the AAAS 1848 Society.

1848 SOCIETY
AAAS

PLEASE CONTACT ME AT:

Name: _____

Address: _____

City: _____ State: _____ Zip code: _____ Country: _____

Email: _____ Phone: _____

RETURN THIS FORM TO:
AAAS Office of Philanthropy and Strategic Partnerships • 1200 New York Avenue, NW • Washington, DC 20005 USA

RESEARCH ARTICLE SUMMARY

PLANT SCIENCE

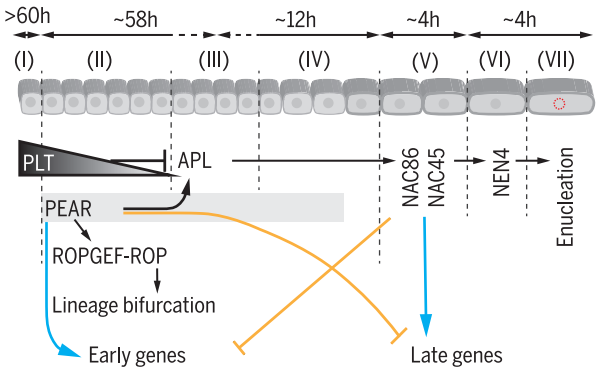
Cell-by-cell dissection of phloem development links a maturation gradient to cell specialization

Pawel Roszak†, Jung-ok Heo†, Bernhard Blob†, Koichi Toyokura†, Yuki Sugiyama‡, Maria Angels de Luis Balaguer‡, Winnie W. Y. Lau‡, Fiona Hamey‡, Jacopo Cirrone‡, Ewelina Madej, Alida M. Bouatta, Xin Wang, Marjorie Guichard, Robertas Ursache, Hugo Tavares, Kevin Verstaen, Jos Wendrich, Charles W. Melnyk, Yoshihisa Oda, Dennis Shasha, Sebastian E. Ahnert, Yvan Saeys, Bert De Rybel, Renze Heidstra, Ben Scheres, Guido Grossmann, Ari Pekka Mähönen, Philipp Denninger, Berthold Göttgens, Rosangela Sozzani*, Kenneth D. Birnbaum*, Yrjö Helariutta*

INTRODUCTION: The plant root grows indeterminately. Continual birth and maturation of cells in a gradient along the root longitudinal axis requires tissue-wide coordination of cell division with cell differentiation. Within the root, a single cell file of developing protophloem is surrounded by other tissues, with each cell type differentiating at its characteristic pace. Despite communication of protophloem cells with the surrounding environment, the developmental program of phloem, within the plant's vasculature, is accelerated compared with certain surrounding cell types. Here, we take advantage of the fast pace of protophloem differentiation and the cellular changes that it undergoes to dissect its developmental trajectory using high-resolution imaging and single-cell omics.

RATIONALE: Single-cell RNA-sequencing (scRNA-seq) analysis as applied to the study of organogenesis typically creates maps of transcriptome activity in various tissue types. Although these approaches characterize gene expression within cells of an organ, the ability to reconstruct the step-by-step changes in cells during maturation is often limited. High-resolution profiling will sample each cellular state along a developmental trajectory and associate each state with developmental changes that lead to cellular specialization. To understand the developmental progression of root phloem cells at a single-cell resolution as related to cellular specializations, we used cell sorting to profile *Arabidopsis thaliana* root tissue and map protophloem-specific transcripts, scRNA-seq to identify molecular transition as cells mature, and live-cell imaging to map molecular states to morphological and cellular events during differentiation.

RESULTS: Long-term live imaging enabled us to determine the duration of the developmental stages and the time one cell spends in each position of the trajectory during protophloem sieve element maturation. We then mapped single-cell transcriptomes corresponding to the 19 cell stages of protophloem development from birth to enucleation. Combining single-cell transcriptomics with cell behavior data from live-imaging experiments, we established seven developmental phases of protophloem development, including early lineage bifurcations, transition from proliferation to differentiation, and, finally, cell enucleation.



Developmental trajectory of protophloem sieve element. Interactions between transcription factors guiding protophloem sieve element development and the length of the identified developmental phases (I to VII). Arrows indicate transcriptional activation. T bars indicate transcriptional inhibition. Colored arrows depict positive and inhibitory interactions identified for early and late factors, respectively, underlying a "seesaw" model. Gray bar indicates PEAR expression domain. Wedge indicates the PLETHORA protein gradient.

The ability to connect cellular development such as lineage bifurcation and enucleation to molecular states using scRNA-seq allowed us to uncover genetic mechanisms that coordinate cellular maturation. First, our analysis revealed the importance of RHO OF PLANTS (ROP) GTPase signaling during early phloem development when the protophloem cell lineage bifurcates to generate metaphloem sieve element and procambium. We found that the expression of the phloem-enriched components of ROP GTPase signaling is triggered by lineage-specific PHLOEM EARLY DNA-BINDING-WITH-ONE-FINGER (PEAR) transcription factors. PEARs also promote phloem differentiation by transcriptional activation of the gene encoding ALTERED PHLOEM DEVELOPMENT (APL), which regulates protophloem sieve element enucleation. In the absence of PEARs, transcription of *APL*, *NAC DOMAIN CONTAINING PROTEIN 45/86* (*NAC45/86*), and *NAC45/86-DEPENDENT EXONUCLEASE-DOMAIN PROTEIN 4* (*NEN4*) is not activated in the protophloem cell lineage and cell enucleation fails.

The genetic cascade, with *PEAR*s handing off late maturation to *APL*, represents a largely autonomous phloem-specific circuit regulating maturation. However, we could also connect the timing of the genetic cascade to broadly expressed master regulators of meristem maturation. Protophloem sieve element differentiation program is temporally coordinated with the rest of the meristem by the broadly acting PLETHORA factors emanating from the stem cell niche. We showed that, although distributed across different tissues, PLETHORA factors directly repress expression of *APL*, counteracting PEARs close to the stem cell niche. The precise timing of developmental mechanisms was critical for proper phloem development; "fail-safe" mechanisms ensured orderly developmental transitions. For example, activation of late genes accompanied repression of early genes of the phloem differentiation program. Ectopic expression of selected late phloem genes in early dividing cells inhibited cell division and promoted cell expansion, two features that characterize late phloem.

CONCLUSION: Using cell sorting, live-microscopy lineage tracing, and transcriptomics, we built a high-resolution blueprint of the genetic program that guides protophloem development. We document even short developmental phases such as cell enucleation, which takes place every 2 hours. Deep, high-resolution single-cell sequencing of the underlying gene-regulatory network revealed a "seesaw" mechanism of reciprocal genetic repression that triggered rapid developmental transitions. Further analysis of this network revealed an interaction of broad versus tissue-specific transcription factors that orchestrates timing of sieve element differentiation. ■

The list of author affiliations is available in the full article online.
***Corresponding author. Email: ross_sozzani@ncsu.edu (R.S.); ken.birnbaum@nyu.edu (K.D.B.); yrjo.helariutta@helsinki.fi (Y.H.)**
†These authors contributed equally to this work.
‡These authors contributed equally to this work.
Cite this article as P. Roszak *et al.*, *Science* 374, eaba5531 (2021). DOI: 10.1126/science.aba5531

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.aba5531

CREDIT: IMAGE BY PAWEL ROSZAK AND BERNHARD BLOB

RESEARCH ARTICLE

PLANT SCIENCE

Cell-by-cell dissection of phloem development links a maturation gradient to cell specialization

Pawel Roszak[1]†, Jung-ok Heo[1,2]†, Bernhard Blob[1]†, Koichi Toyokura[1,3,4,5]†, Yuki Sugiyama[1,6]‡, Maria Angels de Luis Balaguer[7]‡, Winnie W. Y. Lau[8]‡, Fiona Hamey[8]‡, Jacopo Cirrone[9]‡, Ewelina Madej[10], Alida M. Bouatta[11], Xin Wang[2], Marjorie Guichard[12,13], Robertas Ursache[1], Hugo Tavares[1,14], Kevin Verstaen[15,16], Jos Wendrich[17,18], Charles W. Melnyk[19], Yoshihisa Oda[6,20], Dennis Shasha[9], Sebastian E. Ahnert[1,21,22], Yvan Saeys[15,16], Bert De Rybel[17,18], Renze Heidstra[23], Ben Scheres[23,24], Guido Grossmann[12,13], Ari Pekka Mähönen[2], Philipp Denninger[11], Berthold Göttgens[8], Rosangela Sozzani[7]*, Kenneth D. Birnbaum[25]*, Yrjö Helariutta[1,2]*

In the plant meristem, tissue-wide maturation gradients are coordinated with specialized cell networks to establish various developmental phases required for indeterminate growth. Here, we used single-cell transcriptomics to reconstruct the protophloem developmental trajectory from the birth of cell progenitors to terminal differentiation in the *Arabidopsis thaliana* root. PHLOEM EARLY DNA-BINDING-WITH-ONE-FINGER (PEAR) transcription factors mediate lineage bifurcation by activating guanosine triphosphatase signaling and prime a transcriptional differentiation program. This program is initially repressed by a meristem-wide gradient of PLETHORA transcription factors. Only the dissipation of PLETHORA gradient permits activation of the differentiation program that involves mutual inhibition of early versus late meristem regulators. Thus, for phloem development, broad maturation gradients interface with cell-type-specific transcriptional regulators to stage cellular differentiation.

Roots consist of several concentric layers of functionally distinct cell files, which initially bifurcate and establish distinct identities around the quiescent center and its surrounding stem cells. Cells within each file mature through the distinct zones of cell proliferation and differentiation (*1*). For example, in *Arabidopsis thaliana*, the development of the protophloem sieve elements involves a transient period of cell proliferation during which, in addition to amplification of cells within the file, two lineage-bifurcating events take place (Fig. 1A) (*2*). Soon after the cell proliferation ceases, cells of the protophloem sieve element lineage initiate a differentiation process that culminates in enucleation, an irreversible process that gives rise to the mature conductive cells (*3*). Because of specific modulation of the graded distribution of the key phytohormonal cue auxin, the differentiation of protophloem sieve elements occurs faster than that of the other cell files (*4*). Therefore, protophloem sieve element development offers a tractable scheme to understand how the two processes of cell specialization and maturation interact.

Phloem developmental trajectory at single-cell resolution

To understand the process of protophloem sieve element development at a high resolution, we used a combination of approaches based on time-lapse confocal imaging (*5*) and single-cell transcriptomics (*6*). Using the phloem-specific marker *pPEAR1::H2B-YFP pCALS7::H2B-YFP*, we precisely mapped cellular behavior of the 19 cells that constitute the protophloem sieve element developmental trajectory until enucleation, which takes place every 2 hours in the final cell position. The passage of the cell from its "birth" at the stem cell until its enucleation took a minimum of 79 hours (fig. S1 and movies S1 and S2). To dissect the genetic control underlying this temporal progression, we opted for deep profiling of the 19 cells that represent the developmental trajectory of protophloem sieve element using cell sorting and well-based single-cell sequencing over higher-throughput but shallower droplet-based profiling (*6*–*12*). We used fluorescent reporter lines with expression that represents various spatiotemporal domains within the developmental trajectory of the protophloem sieve element (fig. S2, A and B). The single-cell profiles allowed us to cluster cells together with known protophloem sieve element markers to identify 758 cells that densely sampled the 19 cell positions and captured the span of protophloem sieve element maturation (Fig. 1B and fig. S2, C to G).

We sought to use the high-resolution profile of the protophloem sieve element lineage to determine how cell passage through stable signaling gradients in the meristem controls the stages of cellular specialization. In particular, whereas a number of regulators of either phloem cell identity or meristem zonation have been described (*13*, *14*), little is known about how these two regulatory processes interact to control organogenesis. Using Monocle 2 (*15*, *16*), we projected the 758 protophloem sieve element lineage cells into a pseudotemporal order and investigated transcriptional transitions along the developmental trajectory (Fig. 1, B to D). Rather than gradual changes, we observed four transcriptomic domains separated by three narrow transition zones (Fig. 1, D and E, and table S1). On the basis of the alignment with the temporal expression patterns of selected genes, we were able to determine that these domains correspond approximately to cells at positions 1 to 7 [a], 8 to 11 [b], 12 to 15 [c], and 16 to 19 [d], respectively (Fig. 1, D and E, and fig. S3). To further understand which aspects of protophloem sieve element maturation these various positions represent, we extended time-lapse confocal imaging with the more temporally specific marker lines *pNAC86::H2B-YFP* and *pNEN4::H2B-YFP*, which are active at later developmental stages (3). We found that the differentiation time, measured from the last cell division to enucleation, took ~20 hours with some variation up to the final stage defined by expression of *NAC45/86-DEPENDENT*

[1]The Sainsbury Laboratory, University of Cambridge, Cambridge, UK. [2]Institute of Biotechnology, HiLIFE/Organismal and Evolutionary Biology Research Programme, Faculty of Biological and Environmental Sciences, Viikki Plant Science Centre, University of Helsinki, Helsinki, Finland. [3]Department of Biological Sciences, Graduate School of Science, Osaka University, Osaka, Japan. [4]Faculty of Science and Engineering, Konan University, Kobe, Japan. [5]GRA&GREEN Inc., Incubation Facility, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Japan. [6]Department of Gene Function and Phenomics, National Institute of Genetics, Mishima, Japan. [7]Plant and Microbial Biology Department, North Carolina State University, Raleigh, NC, USA. [8]Wellcome Trust and MRC Cambridge Stem Cell Institute and Department of Haematology, University of Cambridge, Cambridge, UK. [9]Computer Science Department, Courant Institute for Mathematical Sciences, New York University, New York, NY, USA. [10]Faculty of Biochemistry, Biophysics and Biotechnology, Jagiellonian University, Kraków, Poland. [11]Plant Systems Biology, Technical University of Munich, Freising, Germany. [12]Institute of Cell and Interaction Biology, CEPLAS, Heinrich-Heine-University Düsseldorf, Düsseldorf, Germany. [13]Centre for Organismal Studies, Heidelberg University, Heidelberg, Germany. [14]Bioinformatics Training Facility, Department of Genetics, University of Cambridge, Cambridge, UK. [15]Data Mining and Modelling for Biomedicine, VIB Center for Inflammation Research, Ghent, Belgium. [16]Department of Applied Mathematics, Computer Science and Statistics, Ghent University, Ghent, Belgium. [17]Department of Plant Biotechnology and Bioinformatics, Ghent University, Ghent, Belgium. [18]VIB Center for Plant Systems Biology, Ghent, Belgium. [19]Department of Plant Biology, Swedish University of Agricultural Sciences, Uppsala, Sweden. [20]Department of Genetics, the Graduate University for Advanced Studies, SOKENDAI, Mishima, Japan. [21]Department of Chemical Engineering and Biotechnology, University of Cambridge, Cambridge, UK. [22]The Alan Turing Institute, British Library, London, UK. [23]Department of Plant Sciences, Wageningen University and Research, Wageningen, Netherlands. [24]Rijk Zwaan R&D, 4793 Fijnaart, Netherlands. [25]Center for Genomics and Systems Biology, New York University, New York, NY, USA.

***Corresponding author. Email: ross_sozzani@ncsu.edu (R.S.); ken.birnbaum@nyu.edu (K.D.B.); yrjo.helariutta@helsinki.fi (Y.H.)**

†These authors contributed equally to this work. ‡These authors contributed equally to this work.

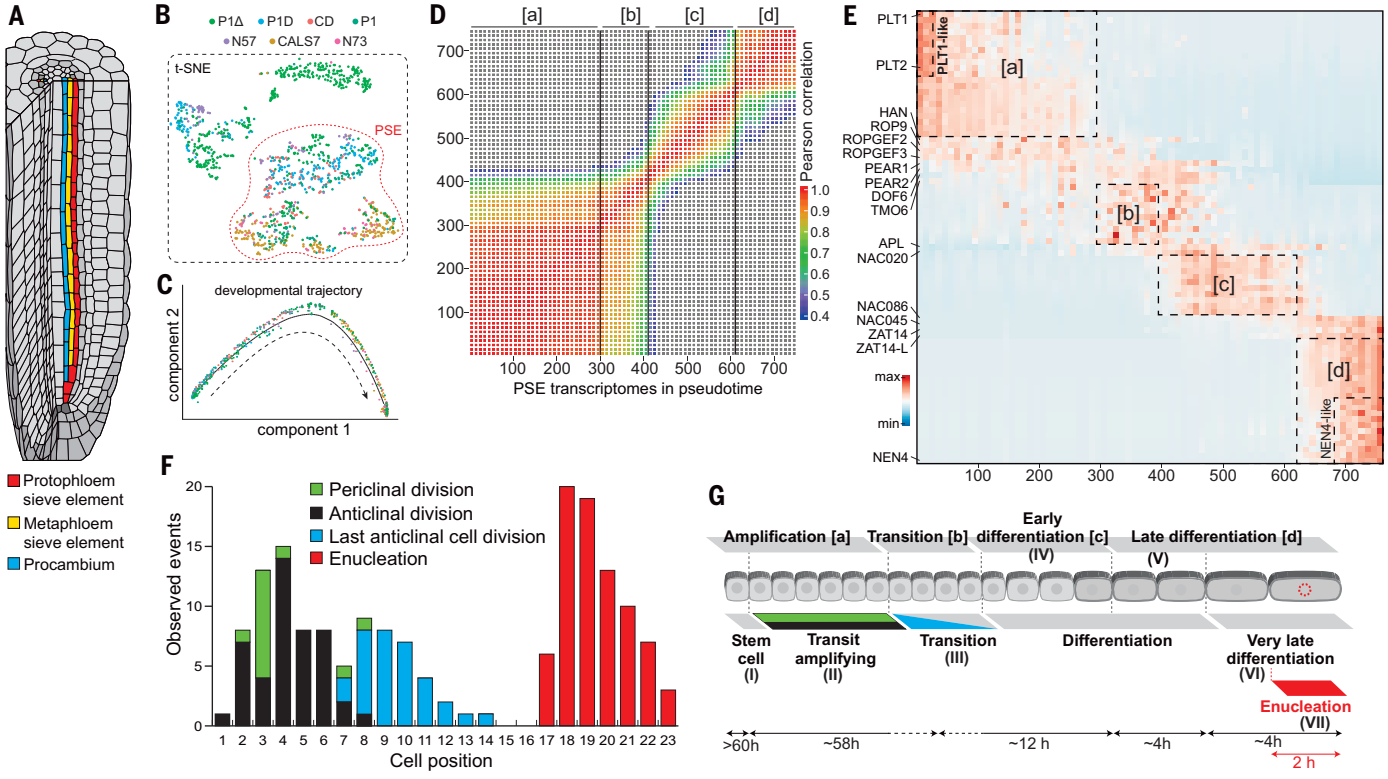


Fig. 1. Phloem development at single-cell resolution. (**A**) Schematic of the *Arabidopsis* root tip depicting the positions of the protophloem sieve element, metaphloem sieve element, and procambium cell lineages originating from a single phloem stem cell. (**B**) t-distributed stochastic neighbor embedding (t-SNE) plot of 1242 transcriptomes of cells sorted with P1Δ, P1D, CD, P1, N57, CALS7, and N73 reporter lines specific to the different domains of the developing phloem. Indicated protophloem sieve element cells were used for the pseudotime trajectory analysis (fig. S2, supplementary materials). (**C**) Protophloem sieve element transcriptomes ordered along developmental trajectory using Monocle 2. (**D**) Heatmap of Pearson correlation along the pseudotime trajectory. Vertical lines indicate the three strongest correlation drops and separate four groups of transcriptomes with higher similarity [a], [b], [c], and [d]. (**E**) Gene expression heatmap of protophloem sieve element regulators and the 10 most specific genes from the four groups defined in (D) and the nested *PLT1* ("PLT1-like") or *NEN4* ("NEN4-like") expression domains in pseudotime-ordered protophloem sieve element transcriptomes. (**F**) Histogram of cell behavior based on long-term live imaging. (**G**) Seven domains and the time cells spent in each position of the developing protophloem sieve element as determined by the transcriptomics (above) and live imaging (below): (I) "stem cell," position 1 [a], t > 60 hours; (II) "transit amplifying," positions 2 to 9 [a], t = 58 hours, SD ± 8.1 hours; (III) "transitioning, positions 8 to 11 [b]; (IV) "early differentiating," positions 10 to 15 [c], t = 12 hours; (V) "late differentiating," positions 16 and 17 [d], t = 4 hours; (VI) "very late differentiating – NEN4-like," positions 18 and 19 [d], t = 4 hours; and (VII) "enucleating," position 19 [d], t = 2 hours (movies S1 and S2).

EXONUCLEASE-DOMAIN PROTEIN 4 (*NEN4*) (active in positions 18 and 19) (Fig. 1E; fig. S1, D, H, and I; and movies S1 to S12). In summary, on the basis of the high congruence of the single-cell transcriptome and live imaging data, we were able to assign seven distinct developmental phases along the protophloem sieve element trajectory: (I) "stem cell," position 1; (II) "transit amplifying," positions 2 to 9; (III) "transitioning," positions 8 to 11; (IV) "early differentiating," positions 10 to 15; (V) "late differentiating," positions 16 and 17; (VI) "very late differentiating," positions 18 and 19; and (VII) "enucleating," position 19 (Fig. 1, F and G; fig. S1; and table S2).

PEARs promote lineage bifurcation through GTPase signaling

Proximal to the stem cell (I) developmental phase, the first distinctive feature of the protophloem sieve element lineage is the bifurcation of the procambial and metaphloem cell files from the progenitor protophloem sieve element lineage through a pair of subsequent periclinal (asymmetric) cell divisions in the domain of transit amplifying cells (II). Using the single-cell lineage and imaging analysis, we sought to precisely map these divisions (Fig. 2A). We observed that the first periclinal division followed exclusively a rare event of phloem stem cell division (movie S13 and fig. S4A). The second, more frequent, periclinal division was observed predominantly at position 3 (Fig. 1F). We have recently shown that the PHLOEM EARLY DNA-BINDING-WITH-ONE-FINGER (PEAR) transcription factors (TFs) (transcribed in domains I to IV) mediate early asymmetric divisions in the phloem lineage and laterally adjacent procambial cells in a cell-autonomous and non-cell-autonomous manner, respectively (*17*).To identify potential downstream effector genes for this PEAR function, we focused on the genes enriched in the expression domain of *pPEAR1Δ::erVenus* marker line (see the materials and methods), capturing the bifurcation events and the resulting protophloem, metaphloem, and procambium cell lineages (Fig. 2B and fig. S4B).

Among the sieve element–enriched genes that were highly expressed in single-cell profiles preceding and during the bifurcation (domain II), we identified and validated the protophloem sieve element abundant expression of Rho-related guanosine triphosphatase (GTPase), Rho of plants 9 (ROP9) (*18*), as well as several genes encoding PRONE-type ROP guanine nucleotide exchange factors (ROPGEF) (Fig. 2, B to D, and fig. S4, B, C, and F) (*19*). ROP GTPase signaling controls the polarity of the multiple cell types during cell differentiation (*20*–*22*) and specific cell division events (*23*–*25*). Subsequently, we determined that ROPGEF3 and ROPGEF5 expression in the protophloem

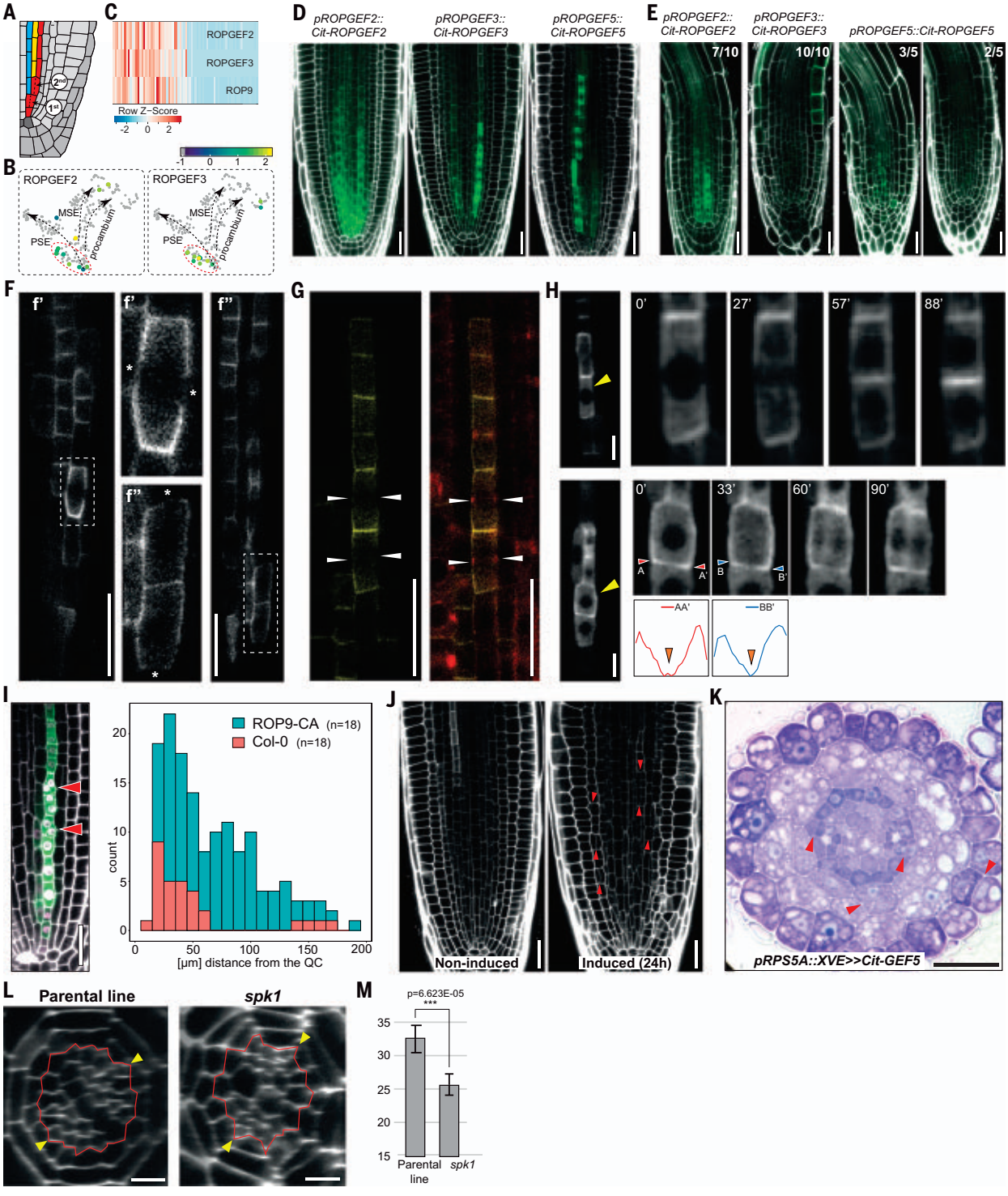


Fig. 2. PEARs control asymmetric divisions by promoting ROP signaling in the phloem pole. (**A**) Schematic indicating position of the two periclinal divisions in the phloem cell lineage. (**B**) Expression of ROPGEF2 and ROPGEF3 at the time of phloem lineage bifurcation. (**C**) Peak expression of ROPGEF2, ROPGEF3, and ROP9 in the early phloem cells as detected in the pseudotime-ordered single-cell protophloem sieve element transcriptome data. (**D**) Expression pattern of phloem-enriched ROPGEFs. ROPGEF3, and ROPGEF5 share similar expression domains, which are enriched in protophloem sieve element and adjacent vascular cell files; ROPGEF2 is expressed in protophloem sieve element but also in other outer procambial cells and pericycle (fig. S4D). Scale bars, 25 µm. (**E**) Expression of ROPGEF2, ROPGEF3, and ROPGEF5 in the *pear* sextuple mutant background. Scale bars, 25 µm. (**F**) Protein localization of *pROPGEF5::Cit-ROPGEF5* during anticlinal (f') and periclinal (f'') cell division. Gaps in ROPGEF5 signal are indicated with an asterisk. Scale bars, 25 µm. (**G**) Depletion of Cit-ROPGEF5 membrane signal at the cortical division zone (CDZ) during cell division. CDZ is marked by accumulating cortical microtubules (mCherry-TUA5) forming a pre-prophase band (white arrowheads). Scale bars, 25 µm. (**H**) Time course analysis of the dynamic pattern of active ROP signaling in the dividing phloem cells. Depletion of *pPEAR1::mScarlet-I-MIDD1ΔN* signal at the CDZ in the anticlinally (upper row) and periclinally (lower row) dividing cells (yellow arrowheads). Quantification of

fluorescence signal intensity in the periclinally dividing cells. Scale bars, 10 μm. (**I**) Quantification of asymmetric cell divisions (red arrowheads) in the protophloem sieve element cell lineage after expression of constitutively active ROP9 (Q64L) (*pPEAR1::XVE>>ROP9^{CA}*). Scale bars, 25 μm. (**J**) Ectopic asymmetric cell divisions (red arrowheads) 24 hours after induction of ectopic Cit-GEF5 expression (*pRPS5A::XVE>>Cit-GEF5*). Scale bars, 25 μm. (**K**) Toluidine blue staining of resin sections of a Cit-GEF5–overexpressing line (*pRPS5A::XVE>>Cit-GEF5*) 24 hours after induction. Red arrowheads indicate ectopic periclinal cell divisions in epidermis, endodermis, and pericycle. Scale bars, 25 μm. (**L**) Identification of *spk1* allele in the mutant screen of *pRPS5A::PEAR1-GR* parental line. Presented are images from noninduced plants. Scale bars, 10 μm. (**M**) Quantification of vascular cell files in the *spk1* mutant and its parental line *pRPS5A::PEAR1-GR*. Neither line was induced, *t* test indicates statistical difference.

sieve element lineage is dependent upon PEAR factors, based on the spatiotemporal correlation as well as the analysis of transcriptional reporters in the *pear* sextuple mutant background (Fig. 2E). In addition, functional analysis of the PEAR-binding sites previously indicated by the DNA affinity purification sequencing (DAP-seq) technique (*26*) in the promoter region of ROPGEF genes affected their expression level (fig. S4D) (17), suggesting a direct interaction.

In the dividing cells, ROPGEFs accumulated broadly at the cell membrane but were depleted from the expected position of the cortical division zone, which demarcates the future division plane (Fig. 2F) (*25*). Indeed, observed gaps in ROPGEF localization coincided with the position of microtubule array called the preprophase band, the earliest marker of the cell division plane in plants (Fig. 2G and fig. S4E) (*25*). ROPGEFs catalyze the disassociation of GDP from an inactive ROP-GDP complex that enables quick binding of free cytosolic GTP and thus activates ROP signaling. In the active state, ROP-GTP interacts with a number of different effector proteins to mediate downstream signaling (*27*). To detect the cellular position of the active ROP signaling in relation to the periclinal and anticlinal cell division planes in phloem, we used a molecular biosensor of ROP signaling that consists of the fluorescently tagged, ROP-GTP–binding domain from the MICROTUBULE DEPLETION DOMAIN1 (MIDD1ΔN) effector protein (*28*). Similar to the localization of ROPGEFs, subcellular localization of active ROP signaling was detected on the cell membrane and was absent in the cortical division zone of protophloem sieve element cells during mitosis (Fig. 2H).

To test whether ROP signaling plays a decisive role in the selection of the cell division plane, we generated an inducible line expressing the constitutively active form of ROP9 (ROP9^{CA}) (see the materials and methods) and lines ectopically expressing phloem-enriched ROPGEFs. Accumulation of ROP9^{CA}-3xYFP on the radial walls of the protophloem sieve element lineage correlated with cell expansion to the radial direction and reorientation of the cell division plane (Fig. 2I and fig. S4F). Ectopic expression of ROPGEFs resulted in ectopic periclinal cell divisions in the outer root layers and pericycle, which rarely undergo such division (Fig. 2, J and K, and fig. S4, G and H). Members of the PRONE-type ROPGEF gene family in *Arabidopsis* have previously been proposed to act redundantly in a number of processes in which they activate ROP signaling (*29*). However, loss of *SPIKE1* (*SPK1*), encoding a single-copy ROP-interacting DOCK family GEF, causes phenotypes mimicking the combinatorial *rop* mutants (*30*–*32*). Therefore, we focused on the loss-of-function alleles of *SPK1*, one of which we identified in the genetic screen for factors promoting formative (periclinal) cell divisions (see the supplementary materials). In the *spk1* loss-of-function mutant, we detected a reduction in periclinal divisions in several tissues, including the protophloem sieve element cell lineage (Fig. 2, L and M, and fig. S4, I to K). We conclude that in the transit amplifying cells (domain II, positions 2 to 9), PEAR function promotes the bifurcation involving the emergence of the protophloem sieve element cell lineage by switching the orientation of the cell divisions at least partially through the activation of ROPGEF-ROP signaling module.

PLETHORAs stage *APL* expression and phloem differentiation

Another distinctive feature of the early protophloem sieve element developmental trajectory is the transition from cell division to cell differentiation (domain II to III to IV). This transition mapped closely to the first major change in the protophloem sieve element transcriptome. In the first transcriptomic domain (domain I to II), we detected transcripts of the PLETHORA gene family (Fig. 1E), the relatively persistent proteins of which are known to spread shootward through cell-to-cell movement. This movement, together with a mitotic dilution effect, contributes to the formation of the shootward protein gradient of PLETHORAs (*14*). Prior work has shown that PLETHORA TFs broadly regulate meristem development, promoting cell division at moderate concentrations, and then permitting elongation and differentiation as levels drop (*14*, *33*, *34*). However, it is not clear how individual cell files interpret the meristem-wide PLETHORA gradient for their own specialized differentiation.

We hypothesized that the PLETHORA gradient might mediate the first transcriptional shift (from domain II to III) toward protophloem sieve element differentiation by permitting a new set of transcripts to be expressed (Fig. 3A). We tested this hypothesis by driving PLETHORA2 (PLT2) under several promoters that extended its expression in the protophloem sieve element in later maturation stages than its native domain (Fig. 3B and fig. S5A). When using the *pNAC86::XVE* inducible promoter, which is active in domains V to VII (*3*, *35*), ectopic PLT2 delayed protophloem sieve element enucleation (Fig. 3B and fig. S5A). Transcriptional profiling of phloem cells expressing the construct showed an up-regulation of genes (table S3) that mapped to the early stages of the protophloem sieve element single-cell trajectory (from domain I to II), the known PLT2 protein gradient (Fig. 3C). These results suggest that extending the PLT2 gradient is sufficient to prolong the early stages of meristem maturation within the protophloem sieve element lineage, providing a connection between the maturation of a specific cell file and a meristem-wide protein gradient. In addition, in the pseudo-time ordered single cells, we could detect complementary oscillatory patterns of the putative S-phase and G_2- to M-phase genes that were also among the up-regulated PLETHORA targets, apparently corresponding to regular progressions through the cell cycle (Fig. 3C and fig. S5B). Furthermore, *ALTERED PHLOEM DEVELOPMENT* (*APL*), *NAC45/86*, and *NEN4*, known key regulators of the protophloem sieve element enucleation pathway (*3*), were among the PLT2–down-regulated genes (fig. S5C and table S3). This is consistent with the presence of *APL* in the large set of genes down-regulated by PLETHORA overexpression (*33*). We validated the down-regulation of *APL* and *NEN4* by ectopic *PLT2* expression with in situ hybridization (Fig. 3D and fig. S5D). We also monitored a shootward shift of *APL* expression domain in the roots after conditional ectopic induction of *PLT2* expression. The induction of PLT2 in the phloem cells beyond its native domain confirmed that activation of APL-dependent genetic program requires dissipation of the PLETHORA gradient (Fig. 3E). To determine the role of PLETHORAs in controlling the transition between transit amplification and differentiation in phloem, we used an inducible, tissue-specific CRISPR/Cas9 approach to mutate *PLT2* specifically in the protophloem sieve element cell file (*36*). We observed an acceleration of the protophloem sieve element differentiation, as well as the expression of *pAPL::erTurq* reporter, toward the quiescent center without affecting the broader meristem size or root growth, showing

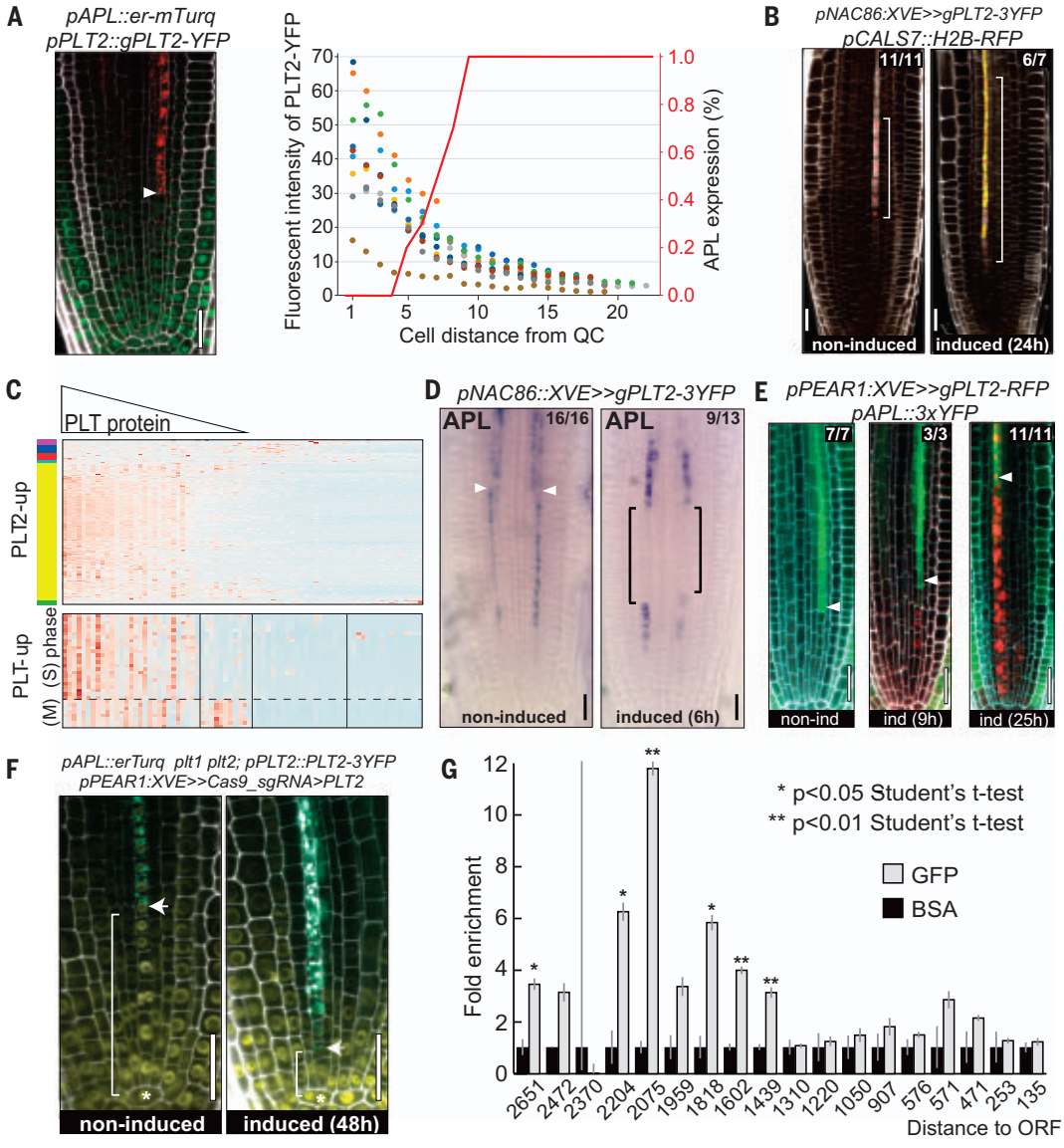


Fig. 3. PLT2 inhibits phloem differentiation by directly repressing APL expression. (**A**) Quantification of the fluorescent intensity of PLT2-YFP in protophloem sieve element cells of nine roots indicated with dots of different colors. Percentage of roots expressing *APL* in a given protophloem sieve element cell is indicated as a red line (n = 9). Onset of *APL* expression coincides with diminishing level of PLT2 protein. Arrowhead indicates the onset of *APL* expression in the protophloem sieve element. (**B**) Ectopic expression of *PLT2* under the *pNAC86::XVE* promoter delays protophloem sieve element enucleation. Square brackets indicate extended expression domain of *pCALS7::H2B-RFP*, a reporter used for monitoring enucleation. (**C**) Native expression profile of PLT2 targets in protophloem sieve element cells ordered in pseudotime. Genes up-regulated after 6 hours of induction of the line shown in (B) are plotted. Upper panel shows gradually diminishing expression of target genes, which reflects the PLT2 protein gradient. Lower panel shows PLT2 up-regulated cell cycle genes with an oscillatory expression pattern. (**D**) In situ hybridization of *APL* before and 6 hours after ectopic expression of PLT2-3xYFP. Arrowheads indicate the positions of protophloem sieve element enucleation beyond which point *APL* is expressed in the phloem pole pericycle, companion cells, and the metaphloem sieve element (fig. S5E). Brackets indicate the *pNAC086* activity domain. (**E**) Time course of transcriptional repression of *APL* in cells ectopically expressing PLT2-RFP under the inducible *pPEAR1::XVE* promoter. (**F**) Early activation of *APL* expression 48 hours after phloem-specific knockout of *PLT2*. (**G**) ChIP-qPCR of PLT2-3xYFP on the *APL* promoter revealing the PLETHORA-binding region −2204 to −1439 bp upstream of the *APL* ORF. All scale bars, 25 μm.

that loss of PLETHORA function in its native domain allows precocious expression of mid- to late-stage protophloem sieve element differentiation regulators (Fig. 3F and fig. S5, F to H).

We sought to further test whether PLT2 directly regulates the protophloem sieve element–specific differentiation program because we found AP2 (a member of the PLETHORA family) family–binding sites in the *APL* promoter region, as determined by the DAP-seq technique (*26*). Indeed, we confirmed the direct binding of PLT2 to several regions of the *APL* promoter by chromatin immunoprecipitation (ChIP) followed by quantitative PCR (qPCR) (Fig. 3G). Furthermore, along with AP2 sites, the *APL* promoter is also enriched for binding sites of HANABA TANARU (HAN), a GATA TF. In turn, *HAN* is a PLETHORA target (*33*) and, accordingly, upon ectopic PLT2 expression we detected *HAN* transcripts expressed in late protophloem sieve element development (fig. S5, C and I). Ectopic HAN expression under *pNAC86:XVE* led to a delay in enucleation (fig. S5J), similar to PLT2 overexpression in the same domain. We conclude that the PLETHORA gradient directly (and possibly in a feedforward manner with HAN) orchestrates protophloem sieve element differentiation by cell autonomously repressing transcription of the phloem regulator *APL*. Overall, our results show how the PLETHORA gradient first promotes cell proliferation in the protophloem sieve element lineage and then helps to time the later stages of cellular maturation.

PEARs promote *APL* to orchestrate phloem differentiation

Given the results above, we reasoned that an early phloem-specific TF must activate *APL* expression. To identify genes that could fill that role, we first generated a list of sieve element genes enriched in our bulk-sorted cells from that tissue compared with published data profiling other tissue types of the root meristem [(*37*); fig. S6A and table S4]. We further narrowed the list by intersecting it with sieve element–enriched genes identified in the cluster analysis of single-cell RNA-sequencing

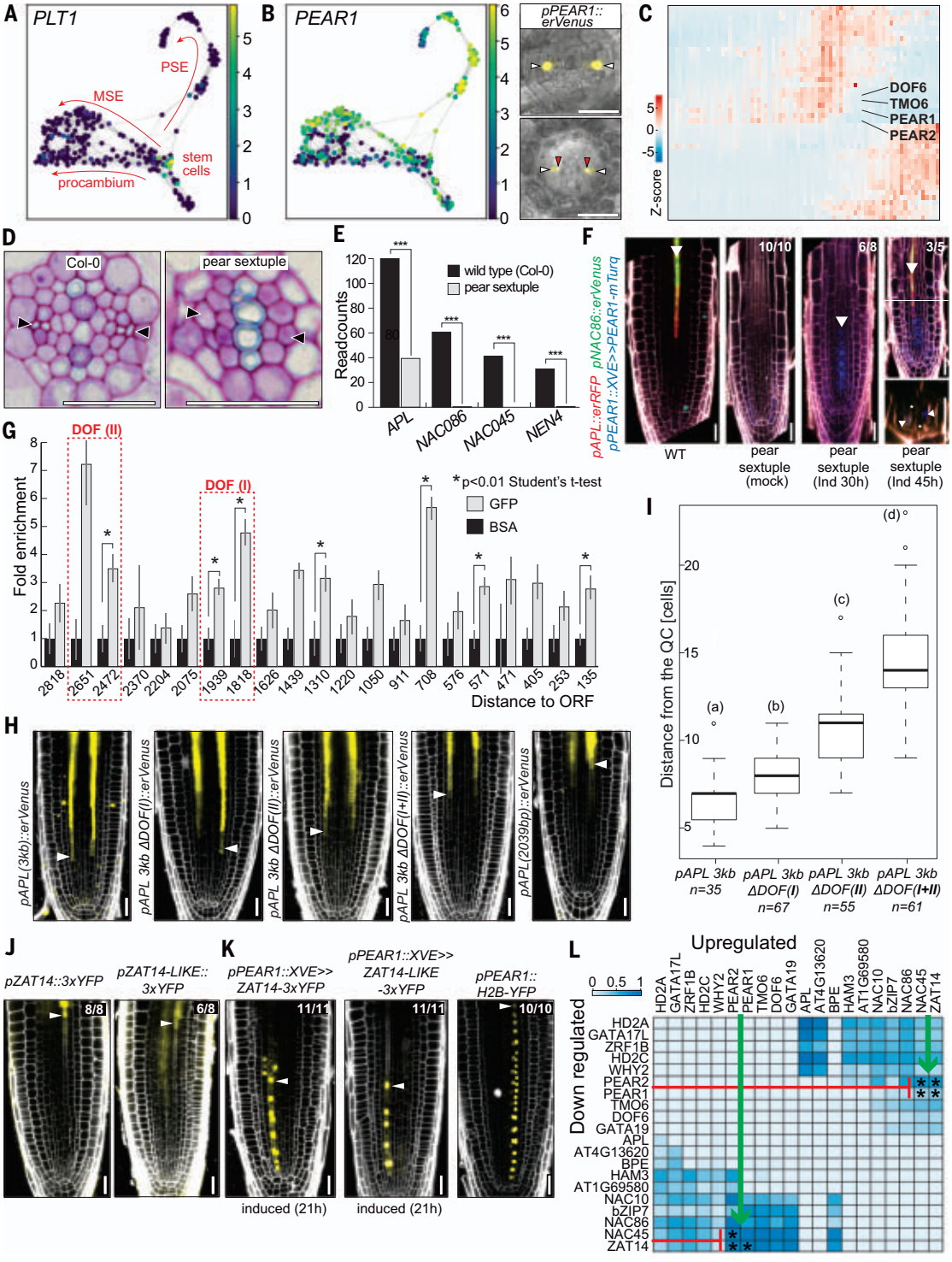


Fig. 4. PEARs orchestrate phloem differentiation. (**A**) Force-directed clustering of 272 single-cell transcriptomes obtained using the *pPEAR1Δ::erVenus* reporter. Plotted is the expression of stem cell–abundant *PLT1*. Arrows indicate cellular trajectories inferred from known gene expression patterns (fig. S6). (**B**) Strong enrichment of *PEAR1* expression in protophloem sieve element and metaphloem sieve element trajectories confirmed by the *pPEAR1::erVenus* reporter line. White arrowheads indicate the protophloem sieve element and red arrowheads the metaphloem sieve element. (**C**) Expression heatmap. PEAR genes are among the earliest phloem-specific TFs. (**D**) Lack of protophloem sieve element differentiation in the mature part of the *pear* sextuple mutant root. Arrowheads indicate the protophloem sieve element position. (**E**) Lack of *APL* pathway activation in the roots of *pear* sextuple mutant based on RNA-seq analysis. (**F**) Inducible expression of *PEAR1-mTurq* is sufficient to activate transcription of *APL* and *NAC86* reporters in the *pear* sextuple mutant background. (**G**) ChIP-qPCR of PEAR1-YFP shows a direct interaction of PEAR1 with *APL* promoter at multiple positions. Two prominent PEAR1-binding sites are indicated with red dashed rectangles. (**H**) Expression patterns of modified *pAPL* reporter lines. Length of the "3kb" promoter equals 2962 bp. DOF(I) and DOF(II) correspond to the two enhancer elements indicated in (G). Details of this modification are provided in fig. S7C. (**I**) Quantification of the onset of *pAPL* expression after

modification of DOF-binding motives. Statistically significant differences between groups were tested using Tukey's post hoc test, $P < 0.05$. Different letters indicate a significant difference at $P < 0.05$. (**J**) Expression of ZAT14 and ZAT14L during late differentiation of protophloem sieve element. Arrowheads indicate the last cell before enucleation. (**K**) Ectopic expression of *ZAT14* and *ZAT14L* under *pPEAR1::XVE* results in cell elongation and inhibition of cell division. Arrowheads indicate the last cell before enucleation. The *pPEAR1::H2B-YFP* line shows the regular number of protophloem sieve element cells. (**L**) Heatmap showing significantly overlapping and oppositely regulated target sets of the 20 most important TFs from the gene-regulatory network model. Color intensity shows the fraction of overlapping target sets. The colormap represents significantly overlapping sets (Fisher's exact test, if $P < 0.05$, value = 1) multiplied by the fraction of overlap. Asterisk indicates experimental validation of up- and down-regulated sets from TF overexpression in vivo (tables S15 and S16). All scale bars, 25 μm.

(scRNAseq) profiles of the *pPEAR1Δ::erVenus* reporter line (table S5; Fig. 4, A and B; and fig. S6, B to H). From this analysis, we identified 542 sieve element–enriched genes (table S6) and corroborated their specificity in the published whole-root scRNAseq atlas (table S7) (*12*). We modeled gene regulation using a machine-learning approach on the pseudotime-ordered 758 single-cell profiles and 4924 highly variable genes. Among the 208 TFs in this dataset, most of the known protophloem sieve element TFs (such as *APL*, *NAC045*, and *NAC086*) were among the top 20 regulators (table S8). We validated the model by comparing predicted targets with genes induced by in vivo ectopic expression of the same TFs, confirming a significant overlap of targets in three of five cases (table S8). Among the top 20 regulators, we also identified four related genes that encode early sieve element–abundant PEAR TFs (*PEAR1*, *PEAR2*, *DNA BINDING WITH ONE FINGER6*, and *TARGET OF MONOPTEROS6*) (Fig. 4C). We recently showed that simultaneous loss of six PEAR genes results in defects in protophloem sieve element differentiation (*17*). We subsequently profiled the transcriptomes of wild-type and *pear* sextuple mutant (Fig. 4D) root meristems and identified 203 down-regulated genes overlapping with our protophloem sieve element–specific gene list (table S9). The expression of *APL*, as well as its downstream targets *NAC045*, *NAC086*, and *NEN4*, was lost in the protophloem tissue of the *pear* sextuple mutant (Fig. 4, E and F, and fig. S7A). Subsequently, expression of *APL* and *NAC086* reporter lines was restored in the pear sextuple mutant upon induction of *PEAR1*, corroborating that transcriptional activation of *APL* in the protophloem sieve element is dependent on the activity of PEAR factors (Fig. 3F).

To test whether PEAR1 can directly regulate the expression of *APL* in its endogenous expression domain (cells 1 to 14), we performed ChIP followed by qPCR using *pPEAR1::PEAR1-GFP* protein fusion and identified multiple PEAR1-binding sites within the *APL* promoter (*pAPL*) (Fig. 4G). Truncation analysis of *pAPL* indicated the presence of an enhancer element responsible for the expression of *APL* in the cells transitioning from cell division to cell differentiation within the 2039- to 2962-bp region upstream of the *APL* open reading frame (ORF) (Fig. 4H). Our ChIP analysis detected a single strong PEAR1-GFP peak in the promoter sequence beyond 2039 bp from the ORF and another strong peak at the upstream end of the 2-kb region, both of which were also detected in the publicly available DAP-seq data (Fig. 4G and fig. S7C) (*26*). Furthermore, within the detected regions (−2672 to −2512 and −1946 to −1844), we identified multiple clusters of DOF-binding motifs (AAAG) (*26*) that constitute an enhancer element required for the transcriptional activation of *APL* in the phloem transition zone (domain III) (Fig. 4, H and I, and fig. S7C). Although the expression of *APL* in the protophloem sieve element is dependent on PEARs (Fig. 4F), *APL* expression domain extends beyond *PEAR* domain (cells 15 to 19; Fig. 1E and fig. S3A). It is possible that either the PEAR proteins and/or *APL* mRNA persist during this period ~10 hours before enucleation. Alternatively, there may be intermediate factors acting downstream of PEARs to promote *APL* expression during late stages of phloem development. Collectively, the data support a role for PEARs in controlling the onset of APL expression to regulate a transition in phloem differentiation. The transition is controlled by the PLETHORAs, which dissipates its own gradient by promoting cell division. When PLETHORA levels decline sufficiently, PEARs can then effectively up-regulate *APL*. The opposing regulation of *APL* by positively regulating PEARS and inhibitory PLETHORAs illustrates how antagonistic mechanisms, one forming a morphogen-like gradient across the meristem, orchestrate developmental timing within a cell file.

Sequential mutual inhibition directs developmental transitioning

The final major transcriptional transition in the phloem lineage occurs between domains IV and V. To explore this transition, we ectopically expressed NEN4 and PLT2 at various developmental stages. When expressed in early ectopic domains, NEN4 expression caused cell death, whereas PLT2 expression forced cells back into the cell cycle. However, later expression of these two TFs had little or no visible effect on cells, showing that the developmental program of domain V appears to be resilient to these perturbations (Fig. 3B and figs. S5A and S9). This indicates that the high number of protophloem sieve element–specific genes during the final 8 hours of differentiation remodel the cellular behavior in an irreversible manner. We next sought to explore how widely the PEARs control transcriptional programs related to this final stage of sieve element development. We combined a gene-regulatory analysis in the *pear* mutant with systematic overexpression and modeling approaches (fig. S7, A and B, and fig. S8). Our analysis revealed that, in addition to the known phloem regulators *APL*, *NAC045*, *NAC086*, and *NAC028*, 10 of 13 newly validated phloem-enriched TFs were dependent on PEARs (fig. S7, A and B, and Fig. 4F). Overexpression of two of these, *ZAT14* (AT5G03510), which was also the third most important TF in the machine-learning model, and its close homolog *ZAT14L* (AT5G04390), led to the arrest of cell cycle and premature cell elongation (Fig. 4, J and K). Transcriptional profiling provided further evidence for a putative dual role in timing cell division and cell expansion (that occurred largely after enucleation in this cell lineage) (tables S10 to S14). In addition, the gene-regulatory network model predicted a pattern of sequential mutual inhibition in the target sets of high-scoring transcriptional regulators (table S15); for example, genes repressed by ZAT14 significantly overlapped with genes activated by the earlier expressed PEARs and vice versa (Fig. 4L). Overexpression analysis confirmed a significant overrepresentation in the overlap between genes up-regulated by PEARs and down-regulated by ZAT14 (table S16) (*17*).

By combining single-cell transcriptomics with live imaging, we have mapped the cellular events from the birth of the phloem cell to its terminal differentiation into phloem sieve element cells spanning a time frame of 79 hours. In the early part of the developmental trajectory, where cells are proliferating, the PEAR factors promote the asymmetric periclinal divisions that result in lineage bifurcation. We pinpoint the ROPGEF-ROP–regulatory module as an effector of early PEAR function in promoting the periclinal cell divisions central to vascular development. In addition, the PEARs activate the final 20-hour terminal differentiation program, which highlights them as central integrators that connect early and late phloem development. Our high-resolution phloem developmental trajectory reveals three abrupt transitions in the gene expression program. The late, PEAR-regulated protophloem sieve element program is directly and antagonistically controlled by the broad PLETHORA

gradient, which connects this morphogen-like gradient to cellular maturation. We propose that mutual inhibition of target genes by sequentially expressed TFs represents a "seesaw" mechanism (fig. S10) that allows rapid transitions and prevents gene expression programs with conflicting effects on cellular physiology (e.g., division versus enucleation). Similar models have been implicated in so-called attractor states in cell fate decisions in animals (*38*). In the future, it will be interesting to determine how conserved these principles of sieve element differentiation are in an evolutionary context, as well as how extensively they apply to other differentiation trajectories in plants.

Methods summary

The single-cell transcriptomic data described herein were generated from the protophloem and metaphloem sieve element, and procambial cells, which were sorted using tissue-specific fluorescent reporter lines. Root tips of 5-day-old *Arabidopsis* plants were used as a tissue material for protoplasting. RNA sequencing of the sorted cells was performed following a well-based Smart-seq protocol. Obtained transcriptomes, corresponding to the cells from the protophloem cell lineage, were ordered in pseudotime using the Monocle 2 package, which generated a single linear protophloem developmental trajectory. Expression profiles and pseudotime coordinates of the known phloem-expressed genes were further confirmed with in situ and reporter line analyses.

The gene-regulatory network was modeled using a random forest machine-learning approach. Selected interactions, representing mutual inhibition (the seesaw model), were confirmed by the transcriptome analysis of lines overexpressing a candidate gene or profiling of the loss-of-function lines.

To understand cell behavior at different developmental phases, confocal long-term live imaging was performed with the protophloem sieve element–specific and nuclear localized reporter line. Up to 5-day-long movies were recorded, and cell behavior, including the number and position of cell divisions, enucleation, and times of these events were recorded.

All details of the methods, including those summarized above, are provided in the supplementary materials.

REFERENCES AND NOTES

1. L. Dolan *et al*., Cellular organisation of the *Arabidopsis thaliana* root. *Development* **119**, 71–84 (1993). doi: 10.1242/dev.119.1.71; pmid: 8275865
2. A. P. Mähönen *et al*., A novel two-component hybrid molecule regulates vascular morphogenesis of the *Arabidopsis* root. *Genes Dev.* **14**, 2938–2943 (2000). doi: 10.1101/gad.189200; pmid: 11114883
3. K. M. Furuta *et al*., Plant development. *Arabidopsis* NAC45/86 direct sieve element morphogenesis culminating in enucleation. *Science* **345**, 933–937 (2014). doi: 10.1126/science.1253736; pmid: 25081480
4. P. Marhava *et al*., A molecular rheostat adjusts auxin flux to promote root protophloem differentiation. *Nature* **558**, 297–300 (2018). doi: 10.1038/s41586-018-0186-z; pmid: 29875411
5. R. Rahni, K. D. Birnbaum, Week-long imaging of cell divisions in the *Arabidopsis* root meristem. *Plant Methods* **15**, 30 (2019). doi: 10.1186/s13007-019-0417-9; pmid: 30988691
6. I. Efroni *et al*., Root Regeneration Triggers an Embryo-like Sequence Guided by Hormonal Interactions. *Cell* **165**, 1721–1733 (2016). doi: 10.1016/j.cell.2016.04.046; pmid: 27212234
7. T. Denyer *et al*., Spatiotemporal developmental trajectories in the *Arabidopsis* root revealed using high-throughput single-cell RNA sequencing. *Dev. Cell* **48**, 840–852.e5 (2019). doi: 10.1016/j.devcel.2019.02.022; pmid: 30913408
8. C. N. Shulse *et al*., High-throughput single-cell transcriptome profiling of plant cell types. *Cell Rep.* **27**, 2241–2247.e4 (2019). doi: 10.1016/j.celrep.2019.04.054; pmid: 31091459
9. K. Jean-Baptiste *et al*., Dynamics of gene expression in single root cells of *Arabidopsis thaliana*. *Plant Cell* **31**, 993–1011 (2019). doi: 10.1105/tpc.18.00785; pmid: 30923229
10. T.-Q. Zhang, Z.-G. Xu, G.-D. Shang, J.-W. Wang, A single-cell RNA sequencing profiles the developmental landscape of *Arabidopsis* root. *Mol. Plant* **12**, 648–660 (2019). doi: 10.1016/j.molp.2019.04.004; pmid: 31004836
11. K. H. Ryu, L. Huang, H. M. Kang, J. Schiefelbein, Single-cell RNA sequencing resolves molecular relationships among individual plant cells. *Plant Physiol.* **179**, 1444–1456 (2019). doi: 10.1104/pp.18.01482; pmid: 30718350
12. J. R. Wendrich *et al*., Vascular transcription factors guide plant epidermal responses to limiting phosphate conditions. *Science* **370**, eaay4970 (2020). doi: 10.1126/science.aay4970; pmid: 32943451
13. M. Bonke, S. Thitamadee, A. P. Mähönen, M.-T. Hauser, Y. Helariutta, APL regulates vascular tissue identity in *Arabidopsis*. *Nature* **426**, 181–186 (2003). doi: 10.1038/nature02100; pmid: 14614507
14. A. P. Mähönen *et al*., PLETHORA gradient formation mechanism separates auxin responses. *Nature* **515**, 125–129 (2014). doi: 10.1038/nature13663; pmid: 25156253
15. C. Trapnell *et al*., The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014). doi: 10.1038/nbt.2859; pmid: 24658644
16. X. Qiu *et al*., Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* **14**, 979–982 (2017). doi: 10.1038/nmeth.4402; pmid: 28825705
17. S. Miyashima *et al*., Mobile PEAR transcription factors integrate positional cues to prime cambial growth. *Nature* **565**, 490–494 (2019). doi: 10.1038/s41586-018-0839-y; pmid: 30626969
18. P. Winge, T. Brembu, R. Kristensen, A. M. Bones, Genetic structure and evolution of RAC-GTPases in *Arabidopsis thaliana*. *Genetics* **156**, 1959–1971 (2000). doi: 10.1093/genetics/156.4.1959; pmid: 11102387
19. A. Berken, C. Thomas, A. Wittinghofer, A new family of RhoGEFs activates the Rop molecular switch in plants. *Nature* **436**, 1176–1180 (2005). doi: 10.1038/nature03883; pmid: 15980860
20. Y. Fu, Y. Gu, Z. Zheng, G. Wasteneys, Z. Yang, *Arabidopsis* interdigitating cell growth requires two antagonistic pathways with opposing action on cell morphogenesis. *Cell* **120**, 687–700 (2005). doi: 10.1016/j.cell.2004.12.026; pmid: 15766531
21. A. J. Molendijk *et al*., *Arabidopsis thaliana* Rop GTPases are localized to tips of root hairs and control polar growth. *EMBO J.* **20**, 2779–2788 (2001). doi: 10.1093/emboj/20.11.2779; pmid: 11387211
22. Y. Oda, H. Fukuda, Initiation of cell wall pattern by a Rho- and microtubule-driven symmetry breaking. *Science* **337**, 1333–1336 (2012). doi: 10.1126/science.1222597; pmid: 22984069
23. J. A. Humphries *et al*., ROP GTPases act with the receptor-like protein PAN1 to polarize asymmetric cell division in maize. *Plant Cell* **23**, 2273–2284 (2011). doi: 10.1105/tpc.111.085597; pmid: 21653193
24. P. Yi, G. Goshima, Rho of Plants GTPases and cytoskeletal elements control nuclear positioning and asymmetric cell division during Physcomitrella patens branching. *Curr. Biol.* **30**, 2860–2868.e3 (2020). doi: 10.1016/j.cub.2020.05.022; pmid: 32470363
25. D. Stöckle *et al*., Putative RopGAPs impact division plane selection and interact with kinesin-12 POK1. *Nat. Plants* **2**, 16120 (2016). doi: 10.1038/nplants.2016.120; pmid: 27501519
26. R. C. O'Malley *et al*.. *Cell* **165**, 1280–1292 (2016). doi: 10.1016/j.cell.2016.04.038; pmid: 27203113
27. G. Feiguelman, Y. Fu, S. Yalovsky, ROP GTPases structure-function and signaling pathways. *Plant Physiol.* **176**, 57–79 (2018). doi: 10.1104/pp.17.01415; pmid: 29150557
28. Y. Sugiyama, M. Wakazaki, K. Toyooka, H. Fukuda, Y. Oda, A novel plasma membrane-anchored protein regulates xylem cell-wall deposition through microtubule-dependent lateral inhibition of rho GTPase domains. *Curr. Biol.* **27**, 2522–2528.e4 (2017). doi: 10.1016/j.cub.2017.06.059; pmid: 28803875
29. P. Denninger *et al*., G. Distinct RopGEFs successively drive polarization and outgrowth of root hairs. *Curr. Biol.* **29**, 1854–1865.e5 (2019). doi: 10.1016/j.cub.2019.04.059; pmid: 31104938
30. J. L. Qiu, R. Jilk, M. D. Marks, D. B. Szymanski, The *Arabidopsis* SPIKE1 gene is required for normal cell shape control and tissue development. *Plant Cell* **14**, 101–118 (2002). doi: 10.1105/tpc.010346; pmid: 11826302
31. D. Basu, J. Le, T. Zakharova, E. L. Mallery, D. B. Szymanski, A SPIKE1 signaling complex controls actin-dependent cell morphogenesis through the heteromeric WAVE and ARP2/3 complexes. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 4044–4049 (2008). doi: 10.1073/pnas.0710294105; pmid: 18308939
32. H. Ren *et al*., SPIKE1 activates ROP GTPase to modulate petal growth and shape. *Plant Physiol.* **172**, 358–371 (2016). doi: 10.1104/pp.16.00788; pmid: 27440754
33. C. Galinha *et al*., PLETHORA proteins as dose-dependent master regulators of *Arabidopsis* root development. *Nature* **449**, 1053–1057 (2007). doi: 10.1038/nature06206; pmid: 17960244
34. L. Santuari *et al*., The PLETHORA gene regulatory network guides growth and cell differentiation in *Arabidopsis* roots. *Plant Cell* **28**, 2937–2951 (2016). doi: 10.1105/tpc.16.00656; pmid: 27920338
35. R. Siligato *et al*., MultiSite Gateway-compatible cell type-specific gene-inducible system for plants. *Plant Physiol.* **170**, 627–641 (2016). doi: 10.1104/pp.15.01246; pmid: 26644504
36. X. Wang *et al*., An inducible genome editing system for plants. *Nat. Plants* **6**, 766–772 (2020). doi: 10.1038/s41477-020-0695-2; pmid: 32601420
37. S. Li, M. Yamada, X. Han, U. Ohler, P. N. Benfey, High-resolution expression map of the *Arabidopsis* root reveals alternative splicing and lincRNA regulation. *Dev. Cell* **39**, 508–522 (2016). doi: 10.1016/j.devcel.2016.10.012; pmid: 27840108
38. J. Shu, H. Deng, Lineage specifiers: New players in the induction of pluripotency. *Genomics Proteomics Bioinformatics* **11**, 259–263 (2013). doi: 10.1016/j.gpb.2013.09.005; pmid: 24095709

ACKNOWLEDGMENTS

We thank Y. Kondo, P. Benfey, and O. Leyser for providing seeds of *pNAC57::2GFP*, *pSHR::erGFP*, and *pPIN4::PIN4-GFP* reporters, respectively; the staff of the Flow Cytometry Core Facility at CIMR for technical support with cell sorting; S. Fujita (National Institute of Genetics, Japan) for providing pDONRP2rP3_MIDD1ΔNcds_stop vector; and R. Wightman and G. Evans for technical support with microscopy experiments. **Funding:** This work was supported by Finnish CoE in Molecular Biology of Primary Producers (Academy of Finland CoE program 2014-2019) decision 271832 (Y.H.); the Gatsby Foundation GAT3395/PR3 (Y.H.); University of Helsinki award 799992091 (Y.H.); ERC Advanced Investigator grant SYMDEV 323052 (Y.H.); NSF-BBSRC MCSB grant 1517058 (R.S. and Y.H.); NSF CAREER MCB grant 1453130 (R.S.); NIH grant GM078279 (K.D.B.); NFS IOS grant 1934388 (K.D.B. and D.S.); NIH grant R35GM136362 (K.D.B.); MRC Clinical Research Infrastructure award MR/M008975/1 (B.G.); core funding from the Wellcome and MRC to the Cambridge Stem Cell Institute (B.G.); Academy of Finland grants 266431 and 307335 (A.P.M. and X.W.); NWO Horizon grant 050-71-054 (R.H.); ERC Starting Grant TORPEDO 714055 (B.D.R.); Research Foundation – Flanders (FWO) Odysseus II GOD0515N (B.D.R.); A Gatsby Foundation CDF grant (S.E.A.); A BOF postdoctoral fellowship from Ghent University (J.R.W.); Wallenberg Academy Fellowship KAW 2016.0274 (C.W.M.); Wellcome Strategic Award 105031/D/14/Z (F.H.); A JSPS Research Fellowship for Young Scientists (K.T.); JSPS KAKENHI grant JP16J00131 (K.T.); a JSPS Overseas Research Fellowship (Y.S.); MEXT KAKENHI grants 16H06280 and 19H05677 (Y.O.); and the Finnish Academy of Science (J.H.). This research was funded, in whole or in part, by the Wellcome Trust (203151/Z/16/Z) and the UKRI Medical Research Council (MC_PC_17230). **Authors contributions:** Conceptualization: P.R., J.H., B.B., K.T., Y.S., M.A.d.L.B., J.C., P.D., G.G., R.H., B.S., C.W.M., B.G., A.P.M., R.S.,

K.D.B., Y.H.; Methodology: J.C., X.W., Y.O., A.M.B., M.G.; Investigation: P.R., J.H., B.B., K.T., Y.S., W.W.Y.L., R.U., E.M.; Data analysis: P.R., J.H., B.B., K.T., Y.S., M.A.d.L.B., F.H., J.C., H.T., K.V., J.W., Y.S., B.D.R., S.E.A., D.S., K.D.B.; Writing: P.R., J.H., B.B., K.D.B., Y.H. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the supplementary materials. Raw RNA-sequencing data for pooled (bulk-sorted) protophloem sieve element cells, single protophloem sieve element cells, and overexpression and mutant profiling have been deposited at the Gene Expression Omnibus of the National Center for Biotechnology Information (submissions GSE142259, GSE140778, and GSE140977, respectively). For the purpose of open access, the author has applied for a CC BY public copyright license to any author-accepted manuscript version arising from this submission.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.aba5531
Materials and Methods
Figs. S1 to S10
Tables S1 to S20
Movies S1 to S13
References (*39*–*53*)
MDAR Reproducibility Checklist

19 December 2019; resubmitted 8 September 2021
Accepted 10 November 2021
10.1126/science.aba5531



Pushing the Boundaries of Knowledge

As AAAS's first multidisciplinary, open access journal, *Science Advances* publishes research that reflects the selectivity of high impact, innovative research you expect from the *Science* family of journals, published in an open access format to serve a vast and growing global audience. Check out the latest findings or learn how to submit your research: **ScienceAdvances.org**

Science Advances
AAAS

GOLD OPEN ACCESS, DIGITAL, AND FREE TO ALL READERS

RESEARCH ARTICLE SUMMARY

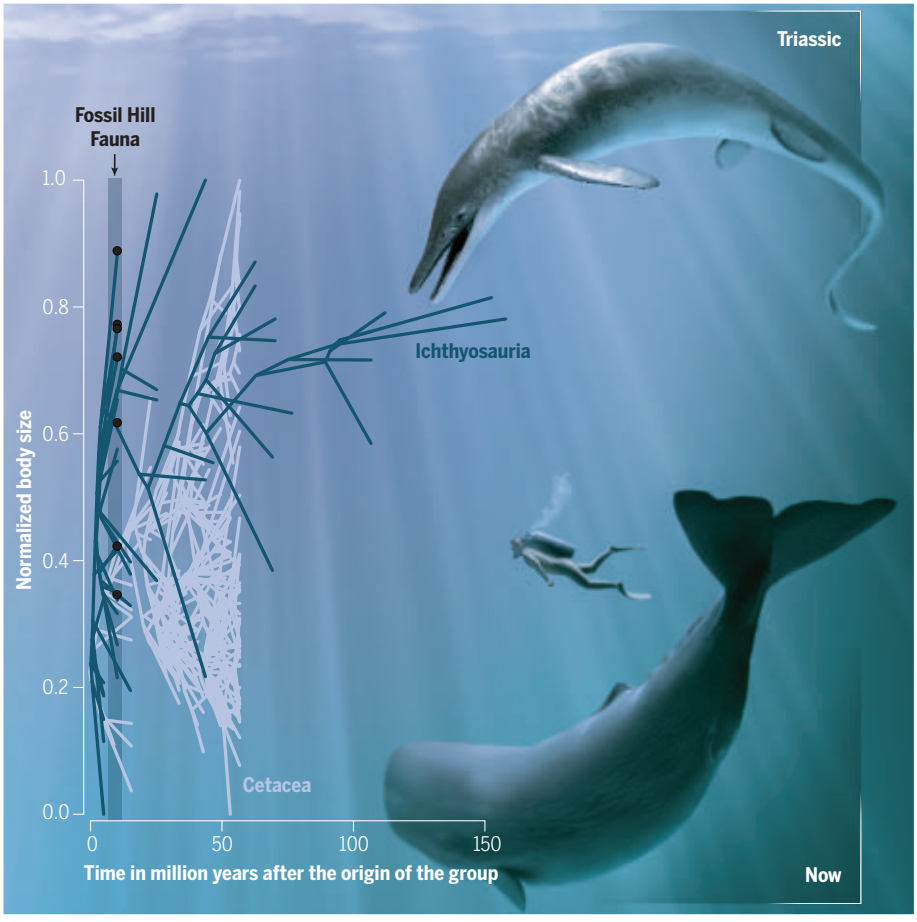
PALEONTOLOGY

Early giant reveals faster evolution of large body size in ichthyosaurs than in cetaceans

P. Martin Sander*, Eva Maria Griebeler*, Nicole Klein, Jorge Velez Juarbe, Tanja Wintrich, Liam J. Revell, Lars Schmitz*

INTRODUCTION: The iterative evolution of secondarily marine tetrapods since the Paleozoic offers the promise of better understanding how the anatomy and ecology of animals change when returning to the sea. Recurring patterns of convergence in the geological past may suggest predictability of evolution when transitioning from full-time life on land to full-time life in the ocean. Ichthyosaurs (fish-shaped marine reptiles of the Mesozoic) and today's cetaceans (whales, dolphins, and porpoises) are two of the most informative lineages to exemplify secondary returns to the sea. The notable resemblance in body shape and lifestyle of ichthyosaurs and cetaceans contrasts with their separation in time by nearly 200 million years, providing an often-cited example of convergent evolution. Ichthyosaurs arose 249 million years ago and populated the oceans for the next 150 million years. Cetaceans did not evolve until about 56 million years ago. As tail-propelled swimmers, ichthyosaurs and cetaceans evolved not only convergent body shapes but also large body sizes.

RATIONALE: The integration of fossil and extant data can improve understanding of aquatic adaptation and gigantism as patterns of convergent evolution, particularly when interpreted in an ecological context. Our paleontological fieldwork in the Fossil Hill Member (Middle Triassic, Nevada, USA) provided the basis for the marine reptile data and resulted in finds of giant ichthyosaurs as part of the pelagic Fossil Hill Fauna. We compiled data for both fossil and living whales from the extensive literature. Together, these data provide the basis for computational analyses of maximum body size and its evolution over time. Modeling of energy flux in the Fossil Hill Fauna helps in understanding how the Fossil Hill ecosystem could have supported several large to giant tetrapod ocean consumers so early in ichthyosaur evolutionary history.



Ichthyosaurs evolved large body sizes earlier in their history than cetaceans. The Fossil Hill Fauna of the Middle Triassic of Nevada, USA, is critical for recognizing this pattern. It features the first ocean giant among tetrapods, only 3 million years after ichthyosaurs first appeared. Whales took comparatively longer to attain similarly large body sizes.

RESULTS: We describe an ichthyosaur with a 2-m-long skull from the Fossil Hill Fauna as a new species of *Cymbospondylus*. At present, this is the largest known tetrapod of its time, on land or in the sea, and is the first in a series of ocean giants. The Fossil Hill Fauna includes several other large-bodied ichthyosaurs in the *Cymbospondylus* radiation. The body-size range in this Triassic fauna rivals the range seen in modern whale faunas, from a total length of about 2 m in *Phalarodon* to more than 17 m in the new species. As preserved in the fossil record, the Fossil Hill Fauna represents a stable trophic network and could even have supported another large ichthyosaur if it bulk fed on small, but abundant, prey such as ammonoids. In absolute time, the new ocean giant lived 246 million years ago, only about 3 million years after the appearance of the first ichthyosaurs. Our research suggests that ichthyosaurs evolved large body size very early on in the clade's history, comparatively earlier than whales.

CONCLUSION: Ichthyosaurs and cetaceans both evolved very large body sizes, yet their respective evolutionary pathways toward gigantism were different. Ichthyosaurs seem to have benefited from the abundance of pelagic conodonts and ammonoids after the recovery from the end-Permian mass extinction, even in the absence of modern primary producers. Cetaceans took different routes, but all appear to be related to trophic specialization, including the loss of teeth in baleen whales (Mysticeti) and the evolution of raptorial feeding and deep diving in toothed whales (Odontoceti). ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: martin.sander@uni-bonn.de (P.M.S.); em.griebeler@uni-mainz.de (E.M.G.); lschmitz@kecksci.claremont.edu (L.S.)
Cite this article as P. M. Sander *et al.*, *Science* **374**, eabf5787 (2021). DOI: 10.1126/science.abf5787

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abf5787

CREDIT: STEPHANIE ABRAMOWICZ

RESEARCH ARTICLE

PALEONTOLOGY

Early giant reveals faster evolution of large body size in ichthyosaurs than in cetaceans

P. Martin Sander[1,2]*, Eva Maria Griebeler[3]*, Nicole Klein[1], Jorge Velez Juarbe[4], Tanja Wintrich[1,5], Liam J. Revell[6,7], Lars Schmitz[2,8]*

Body sizes of marine amniotes span six orders of magnitude, yet the factors that governed the evolution of this diversity are largely unknown. High primary production of modern oceans is considered a prerequisite for the emergence of cetacean giants, but that condition cannot explain gigantism in Triassic ichthyosaurs. We describe the new giant ichthyosaur *Cymbospondylus youngorum* sp. nov. with a 2-meter-long skull from the Middle Triassic Fossil Hill Fauna of Nevada, USA, underscoring rapid size evolution despite the absence of many modern primary producers. Surprisingly, the Fossil Hill Fauna rivaled the composition of modern marine mammal faunas in terms of size range, and energy-flux models suggest that Middle Triassic marine food webs were able to support several large-bodied ichthyosaurs at high trophic levels, shortly after ichthyosaur origins.

Body size is a fundamental attribute of any organism, and extreme body sizes are of special interest to evolutionary biologists. Gigantism is found in different guises in the terrestrial and marine realms (*1*). Several lineages of mammals and reptiles secondarily adapted to marine habitats and diversified to become species-rich clades (*2*), best exemplified by marine mammals since the Paleogene and by marine reptiles of the Mesozoic. Today, multiple species of cetaceans (toothed whales, or odontocetes, and baleen whales, or mysticetes) and pinnipeds (seals and sea lions) inhabit the pelagic ecosystem and differ in body size, feeding strategy, and trophic level (*3*), ranging from macropredatory raptorial feeding (top of the food chain, e.g., killer whales or orcas, *Orcinus orca*) to filter feeding (low in the food chain, e.g., baleen whales). Large marine mammals, especially cetaceans, have been pivotal components of pelagic food webs since at least the late Paleogene, superseding the ichthyosaurs, plesiosaurs, and mosasaurs of the Mesozoic in this role. Body size appears to be a major axis of the phylogenetic and ecological diversification of secondarily marine amniotes.

[1]Abteilung Paläontologie, Institut für Geowissenschaften, Universität Bonn, 53115 Bonn, Germany. [2]The Dinosaur Institute, Natural History Museum of Los Angeles County, Los Angeles, CA 90007, USA. [3]Institut für Organismische und Molekulare Evolutionsbiologie, Evolutionäre Ökologie, Johannes Gutenberg–Universität Mainz, 55099 Mainz, Germany. [4]Department of Mammalogy, Natural History Museum of Los Angeles County, Los Angeles, CA 90007, USA. [5]Anatomisches Institut, Universität Bonn, 53115 Bonn, Germany. [6]Department of Biology, University of Massachusetts Boston, Boston, MA 02125, USA. [7]Universidad Católica de la Santísima Concepción, Concepción, Chile. [8]W.M. Keck Science Department of Claremont McKenna, Scripps, and Pitzer Colleges, Claremont, CA 91711, USA.
*Corresponding author. Email: martin.sander@uni-bonn.de (P.M.S.); em.griebeler@uni-mainz.de (E.M.G.); lschmitz@kecksci.claremont.edu (L.S.)

Analyses of the evolution of body size in independent lineages of pelagic amniotes offer the promise to improve the understanding of the patterns and processes of adaptation to life in marine environments. Repeated transitions from fully terrestrial to obligate marine habitats document how the anatomy and ecology of each lineage evolved in response to the shift from terrestrial to aquatic habitats. Recurring evolutionary patterns may suggest predictability of ecology as well as physiological constraints to maximum and minimum sizes (*4*, *5*). Ichthyosaurs and cetaceans are among the most prominent lineages to exemplify secondary aquatic adaptations. Both clades offer a well-suited model system to understand size evolution in secondary aquatic adaptation and in the sea in general (*2*).

As tail-propelled pelagic tetrapods, ichthyosaurs and cetaceans not only evolved convergent body shapes, lifestyles, physiologies, and feeding strategies as an adaptation to their habitat, but both lineages also evolved after the near-complete collapse of marine ecosystems. However, emerging evidence suggests different trajectories of body-size evolution in the two groups. Mysticetes shifted from gradual evolution of body size to rapid evolution of exceptionally large body sizes late in the history of the clade, concomitant with the extinction of small species. This pattern is presumed to have been facilitated by abundant resources and coastal upwelling (*6*). By contrast, the morphological disparity, the estimated evolutionary rates of discrete characters, and the evolution of skull size of ichthyosaurs all reached an early peak in the Triassic (*7*). The fast increases in disparity measures in early ichthyosaurs reflect rapid lineage diversification and dietary specialization (*8*), including the first aquatic raptorial tetrapod, *Thalattoarchon*, from the early Middle Triassic Fossil Hill Fauna of Nevada, USA (*8*). The presence of an orca-like predator suggested the emergence of food webs that are more similar to modern webs when compared with preceding Paleozoic webs.

However, the giant body sizes of filter-feeding mysticetes and suction-feeding sperm whales (*Physeter macrocephalus*) seemed out of reach for Early and Middle Triassic ichthyosaurs, especially given the absence of environmental indicators of high productivity such as diatoms, autotrophic dinoflagellates, and coccolithophores (*9*). By increasing the amount of energy available for the higher trophic levels of marine ecosystems, the evolution of these relatively large-bodied, planktonic, primary producers is considered to have been a critical precondition for the emergence of modern giants (*3*, *9*). By contrast, small-bodied plankton probably made up the bulk of the primary producers in Triassic oceans, thus limiting the amount of energy available to large-bodied species at higher trophic levels (*3*). By this logic, one would hypothesize that Triassic marine ecosystems should have fewer large species at high trophic levels than modern faunas.

In this contribution, we combine traditional paleontology with computational trait evolution and food web modeling to compare the patterns of body size evolution of ichthyosaurs and cetaceans in an ecological context (Fig. 1). We describe a new ichthyosaur from the early Middle Triassic Fossil Hill Fauna of Nevada, USA (*10*) as *Cymbospondylus youngorum* sp. nov. (Fig. 2) of giant body size (tables S1 and S2) from well-preserved material. The new ichthyosaur lived close to the beginning of Mesozoic marine reptile evolution as part of the recovery from the end-Permian mass extinction (Fig. 3) 252 million years (Ma) ago. The discovery reinforces the emerging pattern of rapid evolution of body size in ichthyosaurs, which, in contrast to cetaceans, must have experienced their most active phase of size evolution in their early evolutionary history, despite the absence of modern primary producers. We infer that the pelagic ecosystems of the early Middle Triassic (244 Ma ago) could, surprisingly, support several large tetrapod ocean consumers.

Systematic Paleontology

Reptilia Linnaeus, 1758 (*11*)
Diapsida Osborn, 1903 (*12*)
Ichthyosauria Blainville, 1835 (*13*)
Cymbospondylus Leidy, 1868 (*14*)

Type species

Cymbospondylus piscosus Leidy, 1868 (*14*)

Referred species

Cymbospondylus petrinus Leidy, 1868 (*14*); *Cymbospondylus buchseri* (*15*); *Cymbospondylus nichollsi* (*16*); *Cymbospondylus duelferi* (*17*); *Cymbospondylus youngorum* sp. nov.

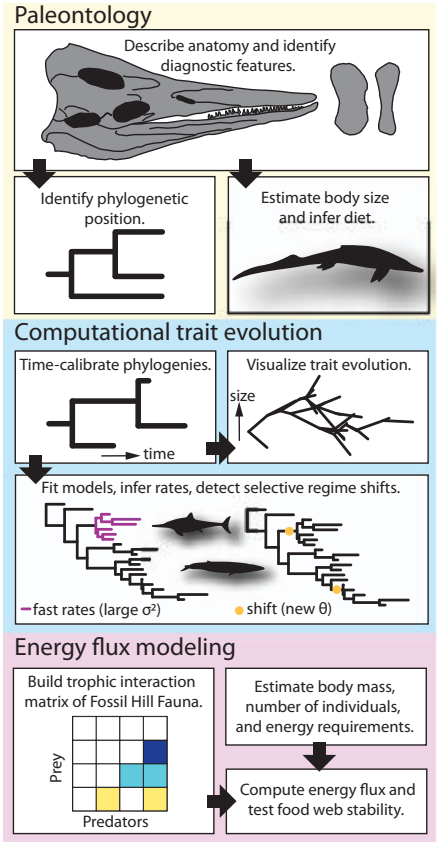


Fig. 1. Conceptual approach of our integrated study. We combine traditional paleontology with computational trait evolution and energy-flux modeling to study macroevolutionary patterns of body size evolution in marine amniotes.

Genus diagnosis

For a recent detailed diagnosis of the genus, see (*17*). In addition, the results of the phylogenetic analysis in this study offer an apomorphy-based diagnosis.

Cymbospondylus youngorum sp. nov. (Fig. 2 and figs. S1 to S5)

Etymology

The species is named in honor of Tom and Bonda Young.

Holotype and only specimen

LACM DI 157871 is the holotype and, as of now, the only recognized individual. LACM DI 157871 is largely articulated and complete from the anterior of the trunk region to the skull, preserved ventral side up. The cervical column back to the middle dorsal vertebrae is present with ribs in articulation. The shoulder girdle is articulated, and the two humeri are situated close to their respective glenoid. At present, the skull (Fig. 2, A to F; figs. S1 and S2; and table S1), the right humerus (Fig. 2, G to K), parts of the shoulder girdle (fig. S3), and some vertebrae are prepared (fig. S3)

Horizon and locality

LACM DI 157871 comes from the Anisian age Fossil Hill Member of the Favret Formation at Favret Canyon, Augusta Mountains, Pershing County, Nevada, USA. The type locality, LACM 8025, is on the northern slope of the rear of Favret Canyon at an elevation of 1676 m. Exact coordinates are on file at the repository. LACM DI 157871 originates from the same general level as the holotype of the macropredatory ichthyosaur *Thalattoarchon saurophagis* (*8*) and the pistosaur *Augustasaurus hagdorni* (*10*) in the lower third of the Fossil Hill member and pertains to the middle Anisian *Taylori* Zone (*18*).

Diagnosis

C. youngorum sp. nov. is diagnosed by a unique combination of the following eight characters (see data S1 for character descriptions and data S3 for a list of synapomorphies): squamosal, participates in supratemporal fenestra (character 72, state 0; Fig. 2); dentary, labial shelf present (character 117, state 1; fig. S2C); angular, extent of anterior lateral exposure is extensive, at least as high and anteriorly as the surangular's exposure (character 120, state 1; Fig. 2); angular, extent of posterior lateral exposure is extensive, with surangular exposure reduced to a thin strip on the lateral surface of the retroarticular process (character 121, state 1; Fig. 2); lower jaw glenoid, deeply excavated and present (character 126, state 1; Fig. 2); dentition, definition of the base of the enamel layer is well defined and precise (character 147, state 1; fig. S2, E and F); humerus, anterior flange is absent (character 200, state 0; Fig. 2 and fig. S6); and humerus, relative anteroposterior width in dorsal view, excluding dorsal and ventral processes, is approximately equal or the proximal end is wider than the distal end (character 206, state 1; Fig. 2 and fig. S6).

The new taxon is characterized by the following autapomorphies: a thick base of bone of attachment of the teeth (Fig. 2 and fig. S2, C and E), the distinctive shape of the scapula with a very large and wide dorsal blade and a narrow ventral part (figs. S3 and S6), the distinctive humerus morphology with a wider proximal than distal end, and a triangular proximal head and triangular shaft cross section (Fig. 2 and fig. S6). Note that these autapomorphies were not added as characters to our character matrix. A differential diagnosis and detailed anatomical descriptions (figs. S2 and S3) and comparisons (figs. S4 to S6 and tables S2 and S3) are provided in the supplementary materials (*10*).

Phylogenetic position

Phylogenetic analyses (*10*) (table S4) indicate that *C. youngorum* sp. nov. is nested within a clade of closely related *Cymbospondylus* species that account for much of the lineage diversity and morphological disparity of large-bodied Early and Middle Triassic ichthyosaurs (Fig. 3 and fig. S7). The close relationship of these species points to an adaptive radiation (as much as one can be recognized in the fossil record of Mesozoic reptiles). Further evidence for such a radiation is that there are no other four seemingly sympatric species of any ichthyosaur genus in the ichthyosaur record and that other finds of *Cymbospondylus* from the late Early and early Middle Triassic are widely distributed across the Northern Hemisphere (*16*).

Our analysis with TNT (a software for phylogenetic analysis) and its "new technology" search algorithm resulted in a tree length of 1225 steps. The four most parsimonious trees were retained (table S4), and the nearly fully resolved strict consensus of the four trees is shown in Fig. 3 and fig. S7. The consistency index of this tree is 0.259, coupled with a retention index of 0.627. The absolute Bremer support of the nodes varies from 1 to 5 (fig. S7). Additional analyses (table S4) confirmed the placement of LACM DI 157871 in a clade with other cymbospondylids, yet the position of the *Cymbospondylus* clade varies with the selection of taxa that were included in the analyses (*10*). We note that the interrelationships of ichthyosaurs remain difficult to resolve, both in the Triassic and the Jurassic part of the tree. This uncertainty reflects the difficulty in resolving ichthyosaur interrelationships in general (*19*) and the need for a concerted effort of redefining and rescoring characters.

Inferred diet and estimated body size

The conical, bluntly pointed tooth crowns of *C. youngorum* sp. nov., in conjunction with the elongate snout, suggest a generalist diet of fish and squid (*20*), as inferred for most ichthyosaurs from teeth and stomach contents (*21*). Considering its size, *C. youngorum* sp. nov. could also have preyed on smaller and juvenile marine reptiles (*10*). The right lower jaw of *C. youngorum* sp. nov. measures 1970 mm from the tip of the dentary to the end of the retroarticular process (table S1). At a total length of 1890 mm (table S1), the skull of LACM DI 15787 is one of the largest complete ichthyosaur skulls known. Although post-Triassic ichthyosaurs never reached the size of Triassic ones again (*7*), there are skulls of *Temnodontosaurus* from the Lower Jurassic of England and Germany (*22*) that are the same length as that of LACM DI 15787 within the limits of preservation. However, these large ichthyosaurs probably were less than 9 m long, having proportionally larger skulls (*22*). Larger skulls than those of LACM DI 15787 and these largest *Temnodontosaurus* specimens are only known from Late Triassic ichthyosaurs, specifically *Shonisaurus popularis* and *Shastasaurus sikanniensis*, with estimated skull lengths of 2750 and 3000 mm, respectively (*7*).

Humerus length is another commonly used proxy for ichthyosaurian body size [(*16*, *17*, *23*);

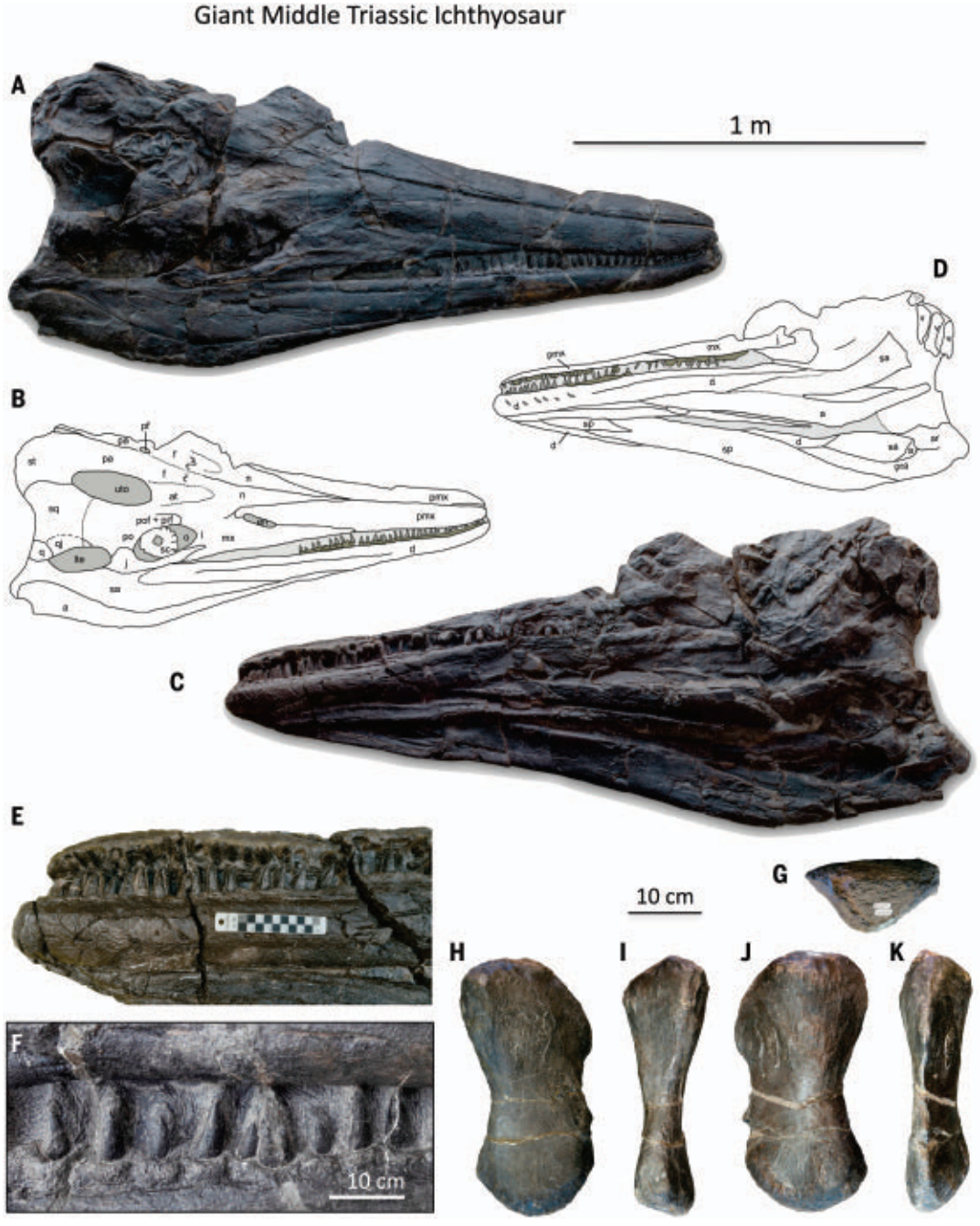


Fig. 2. Skull of the holotype of *C. youngorum* sp. nov. LACM DI 157871. (**A**) Skull in right dorsolateral view. (**B**) Skull sutures. (**C**) Skull in left ventrolateral view. (**D**) Skull sutures. (**E**) Snout in left ventrolateral view. (**F**) Middle part of dentary tooth row in right dorsolateral view. Note the bone of attachment. (**G** to **K**) Right humerus in proximal (G), dorsal (H), posterior (I), ventral (J), and anterior view (K). a, angular; ar, articular; at, anterior terrace; d, dentary; en, external nares; f, frontal; j, jugal; l, lacrimal; lte, lower temporal embayment; mx, maxilla; n, nasal; o, orbita; pa, parietal; pf, parietal foramen; pmx, premaxilla; po, postorbital; pra, prearticular; prf, prefrontal; q, quadrate; qj, quadratojugal; sa, surangular; sc, scleral ring; sq, squamosal; st, supratemporal; uto, upper temporal opening; v, cervical vertebra.

fig. S8A], even though it is available for fewer ichthyosaur taxa than skull length. At 453 mm, the humerus of the holotype of *C. youngorum* sp. nov. is the second largest ichthyosaurian humerus known, translating into a total length of 17.65 m (*10*). The lower 95% prediction interval of its length is 12.48 m; the upper 95% prediction interval is 24.96 m (*10*) (fig. S8A). We also estimated body mass based on a recent dataset for ichthyosaurs (*10*, *24*) (table S5). The regression function (fig. S8B) yielded a body mass estimate of 44,699 kg for the 17.65-m-sized LACM DI 15787 specimen (table S5). The lower mass based on the 95% prediction interval of both regressions is 14,712 kg (lower limit of the 95% prediction interval of body mass against length evaluated for 12.48 m); the upper mass is 135,809 kg (upper limit of the 95% prediction interval of body mass against length evaluated for 24.96 m). These body size metrics of *C. youngorum* sp. nov. result in one of the highest length and mass estimates for any ichthyosaur and the largest taxon of the Middle Triassic: Based on length and scaled-up mass data in Gutarra *et al.* (*24*), *S. popularis* from the late Carnian (about 230 Ma ago) was 13.5 m long and had a mass of 21,651 kg, whereas the middle Norian (about 212 Ma ago) *S. sikanniensis* was 21 m long and had an estimated body mass of 81,497 kg. Taken together, these length and mass estimates place ichthyosaurs in the range of body sizes of living cetaceans, but it appears that ichthyosaurs reached their largest sizes much earlier in clade history than whales.

The fossil record of body-size evolution

Despite the recent interest in ocean gigantism (*4*, *6*, *23*, *24*), detailed comparison of the evolutionary paths to giant body size in cetaceans and ichthyosaurs is lacking so far, as is a phylogenetically comprehensive analysis of

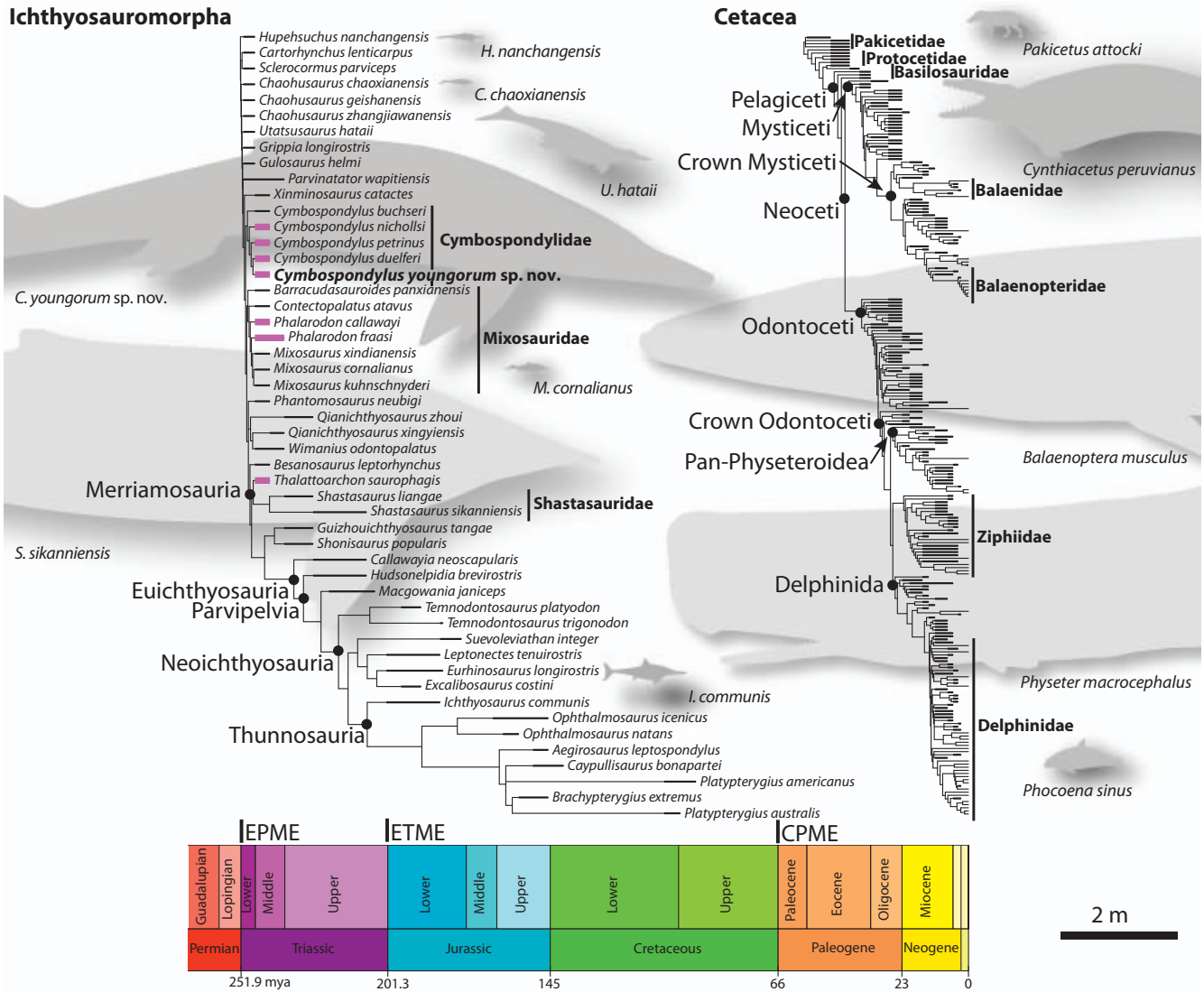


Fig. 3. Time-calibrated phylogenies and body-size illustrations of Ichthyosauria and Cetacea and the relationships of the new giant ichthyosaur *C. youngorum* sp. nov. Ichthyosaurs originated in the late Early Triassic shortly after the end-Permian mass extinction (EPME), survived the end-Triassic mass extinction (ETME), and went extinct in the early Late Cretaceous. Lilac stratigraphic ranges denote taxa from the Fossil Hill Fauna. Cetaceans originated in the late Paleocene after the Cretaceous-Paleogene mass extinction (CPME). See (*10*) for sources of phylogenies and table S6 for image credits. mya, million years ago.

body-size evolution in all Cetacea, including stem taxa (fig. S9). Both ichthyosaurs and cetaceans evolved after mass extinction events (Ichthyosauria after the end-Permian mass extinction event and Cetacea after the Cretaceous-Paleogene mass extinction; Fig. 3), with their respective terrestrial ancestors invading marine ecosystems that differed radically from those that existed before each mass extinction.

The early-branching members of both ichthyosaurs and cetaceans are small compared to later taxa, and both show aquatic adaptations, but to various degrees. The ichthyosaur *Cartorhynchus lenticarpus* from the Early Triassic had a skull length of 55 mm (*7*, *25*), whereas the skull of the cetacean *Pakicetus attocki* from the Eocene reached a width of 127 mm (*10*, *26*). The limbs of *C. lenticarpus* have been interpreted to allow limited terrestrial locomotion, where the cartilage-rich forelimb may have functioned analogously to seal flippers (*25*). The short trunk and snout may have been beneficial for terrestrial locomotion, whereas pachyostotic ribs added to the weight of the animals and perhaps helped with navigating surging water near shores (*25*). *Pakicetus* is interpreted to have inhabited freshwater systems, wading and swimming in shallow water similar to extant hippos, despite its terrestrially adapted limbs (*27*). Upward facing eyes are reminiscent of a crocodilian-like lifestyle at the water surface, enabling vision through air while being submerged in water [analogous to *Tiktaalik* (*28*)]. Increased bone mass through osteosclerosis, stable isotopes, and preferred diet inferred from tooth microwear (*27*, *29*, *30*) further corroborate an aquatic lifestyle.

A literal reading of the geologic time scale implies that body size evolved considerably faster in ichthyosaurs than in cetaceans. Fossils document a rapid size increase in ichthyosaurs from *Cartorhynchus* (55-mm skull length) in the Olenekian at 248.5 Ma ago (*25*) to *C. youngorum* sp. nov. (1890-mm skull length) a mere 2.5 Ma later. Fossils record slower evolution of body size in cetaceans, for example, from *P. attocki* (127-mm skull width) in the late Ypresian (*10*, *26*) to *Basilosaurus isis* (600-mm skull width) in the Priabonian [(*10*) and data S5] 10 to 14 Ma later, and in odontocetes, from *Simocetus rayi* (238-mm skull width) in the Rupelian [(*10*) and data S5] to *Livyathan melvillei* (1970-mm skull width) in the Tortonian [(*10*)

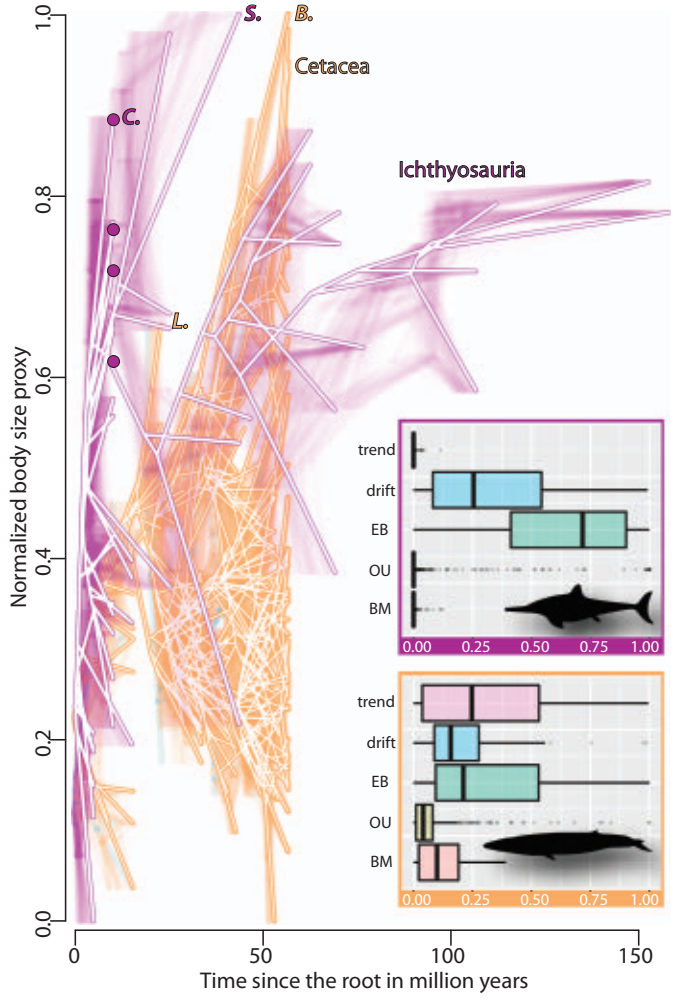


Fig. 4. Body-size evolution in ichthyosaurs and cetaceans compared. Traitgram of body size, normalized such that 0 corresponds to the smallest body size in each group and 1 to the largest for ichthyosaurs (lilac; based on an early-burst model) and cetaceans (ochre; based on a Brownian motion model) (see Methods). Lilac dots indicate *Cymbospondylus* species from the Fossil Hill Fauna. B., *Balaenoptera musculus*, blue whale; C., *C. youngorum* sp. nov.; L., *Llanocetus denticrenatus*, early giant baleen whale; S., *S. sikanniensis*, the largest named ichthyosaur. The inset shows model-fitting results expressed as Akaike weights for five different evolutionary models from 1000 iterations. Boxes represent the interquartile ranges (IQRs), with whiskers extending 1.5 times the IQR outside the boxes. Vertical lines inside the boxes show the median. BM, Brownian motion; EB, early burst; OU, Ornstein-Uhlenbeck.

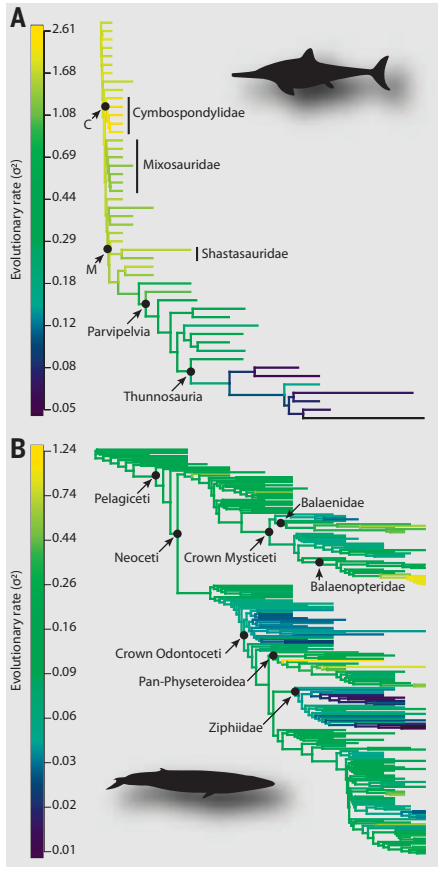


Fig. 5. Exploration of the rate heterogeneity of body-size evolution. (**A** and **B**) Ichthyosaurs (A) feature the fastest rates of body evolution early in their history, whereas cetaceans (B) show a more complex pattern, with fastest rates occurring in later stages of their history. C, Cymbospondylidae; M, Merriamosauria.

and data S5] more than 25 Ma later. These examples illustrate the short time spans of size increases in ichthyosaurs compared with those of cetaceans, but does this pattern hold when analyzed in the context of time-calibrated phylogenies?

Computational modeling of body-size evolution

We first compared the evolutionary patterns of body-size evolution through traitgrams that account for the uncertainty tied to the time calibration of the phylogenies (*10*). These evolutionary traitgrams of ichthyosaur and cetacean body sizes, normalized to maximum body size to enable a better comparison of the pattern, reinforce the notion of fast body-size evolution in ichthyosaurs (Fig. 4). The maximum body size of ichthyosaurs increased dramatically early in their history, whereas cetacean maximum body size increased up to the present day. In our clade-wide model-fitting approaches, the early-burst and trend models are strongly preferred over all other evolutionary models for ichthyosaurs [Fig. 4 and fig. S10; see (*10*)]. By contrast, no strong preference is indicated for any particular model in cetaceans (Fig. 4). Resampling of the cetacean dataset to adjust for differences in sample size, trophic specialization, and the mixture of fossil and extant data did not reveal strong support for an early-burst model (see fig. S11). When restricting the cetacean dataset to the Pelagiceti to account for the differing degrees of aquatic adaptation in the early-branching lineages, an early-burst model is slightly preferred over the others tested (see fig. S11), but the pattern is not nearly as strong as for ichthyosaurs.

To explore the evolutionary patterns at smaller phylogenetic scales and therefore avoid problems related to the impact of taxon selection, we assessed the heterogeneity of the rate of body-size evolution across both the ichthyosaur and cetacean trees (Figs. 5 and 6 and fig. S12). Models that allow for rate heterogeneity can identify specific regions of the tree that signify evolutionary changes and thus circumvent the problem of a clade-wide approach that would ignore the various degrees of aquatic adaptations in the early ichthyosaurs and cetaceans.

Results from a variable-rate model based on Brownian motion, which uses penalized likelihood to estimate the evolutionary rates along all branches in the tree (*31*), support fast evolution of body size in early ichthyosaurs (Fig. 5A). In congruence with observations from the fossil record and clade-wide model fitting, the highest evolutionary rates are found in the

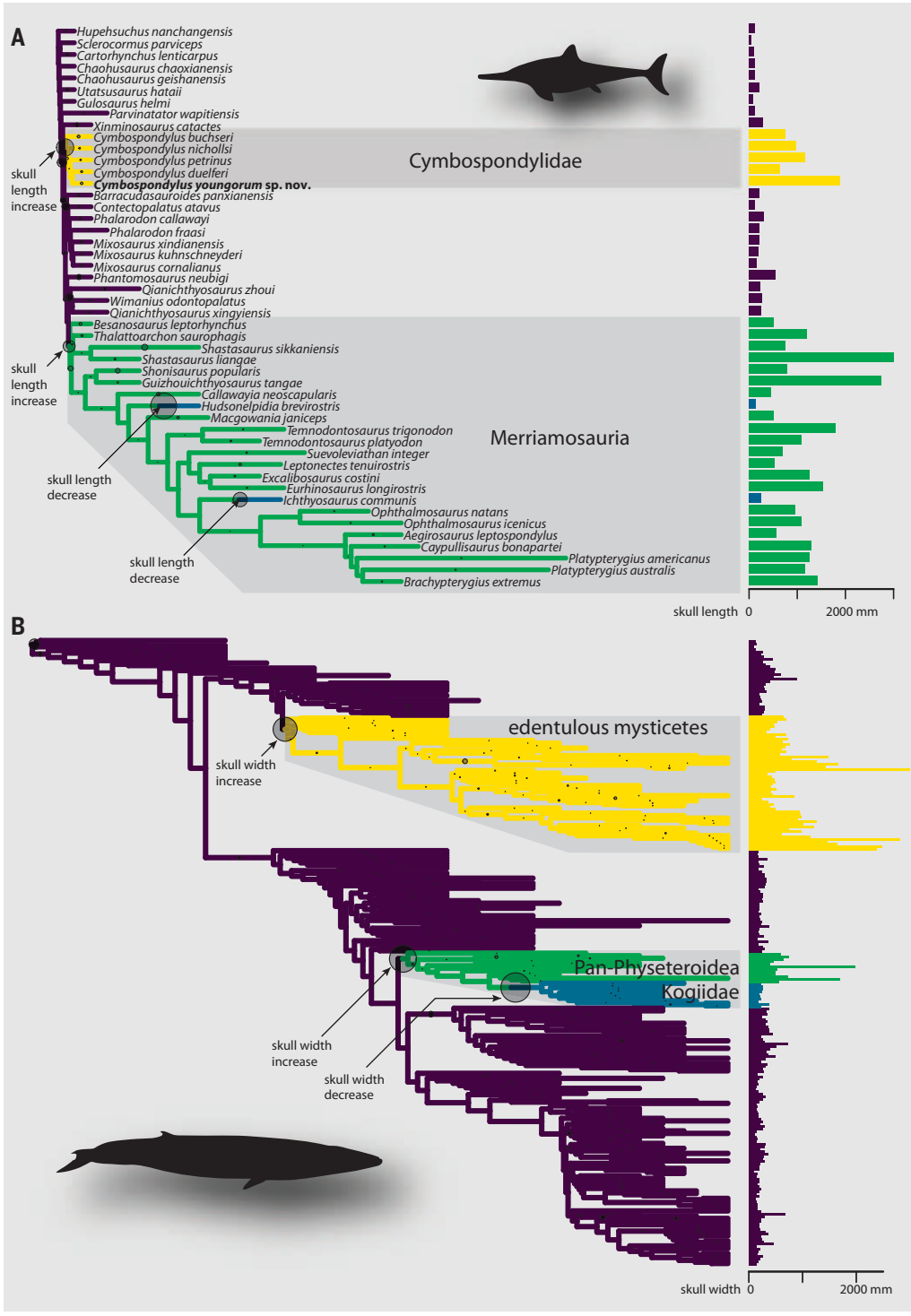


Fig. 6. The adaptive landscape of body-size evolution. (**A** and **B**) Ichthyosaurs (A) feature two early adaptive shifts toward larger skull length, whereas cetaceans (B) entered selective regimes that promoted larger skull width much later in their evolution.

deep regions of the ichthyosaur tree, with some indication that fastest size evolution may have occurred in cymbospondylids and the early phase of merriamosaurs (Fig. 5A). After the initial burst, the estimated evolutionary rates slowed over time.

The distribution of evolutionary rates in the cetacean tree forms a stark contrast to the ichthyosaur pattern (Fig. 5B). The deep parts of the cetacean tree feature average rates. Surprisingly, the body size of both early crown mysticetes and odontocetes evolved slowly, before accelerating in two clades of mysticetes with living representatives, the Balaenidae (right and bowhead whales) and Balaenopteridae (rorquals), as found in a previous study (*6*). Among odontocetes, the fossil lineages of the Pan-Physeteroidea, in particular, a clade that contains living sperm whales, are characterized by fast rates, whereas Ziphiidae (beaked whales) have comparatively slow rates of size evolution (Fig. 5B).

Whereas rate heterogeneity models can identify phases of fast and slow morphological evolution, the determination of the direction of evolution, that is, evolution toward larger and smaller body sizes in specific lineages,

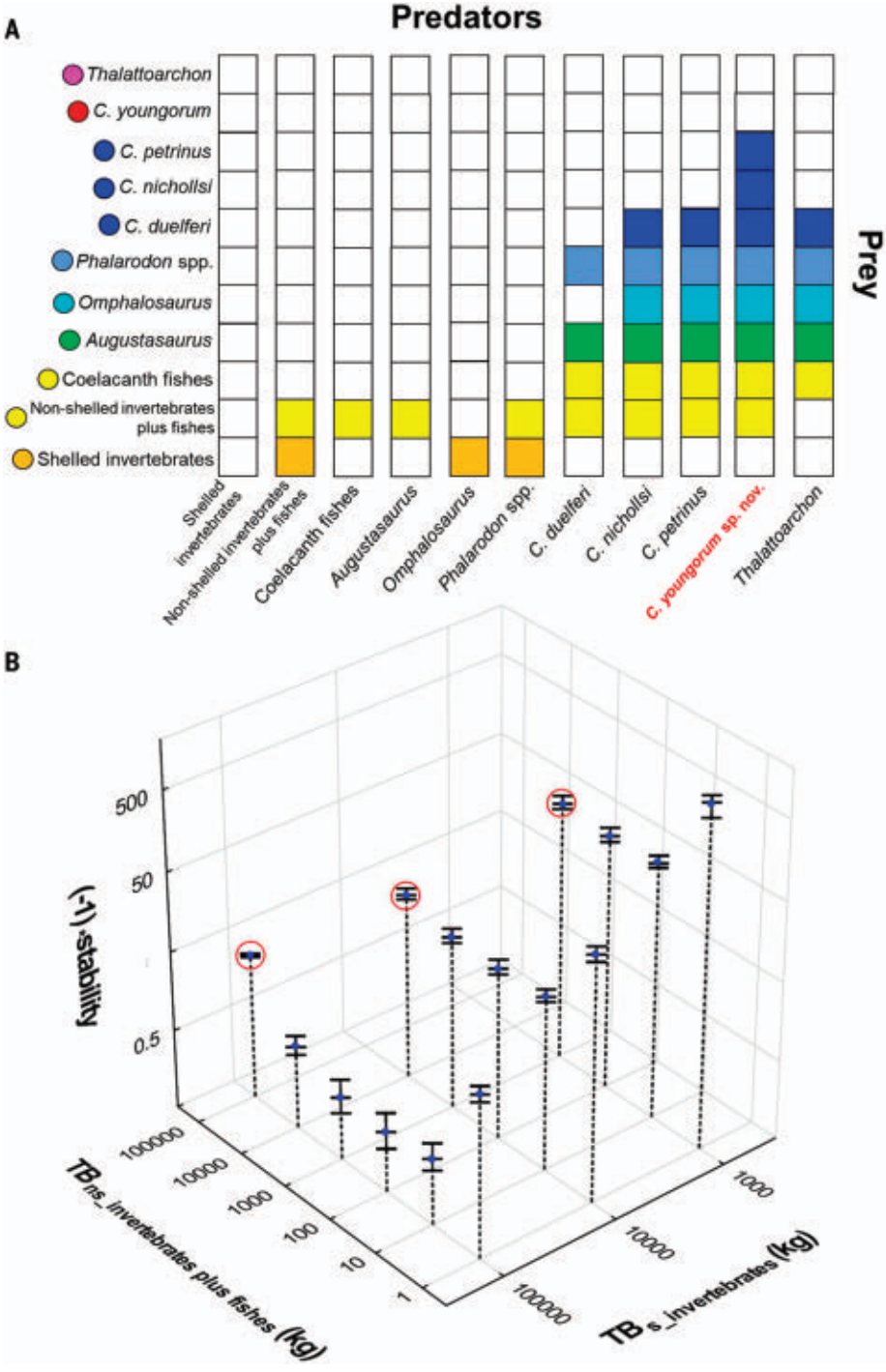


Fig. 7. Energy-flux model and food web stability. We tested whether the food web as preserved is functional and stable over ecological time. All shelled invertebrates (specifically ammonoids), fish, and amniote taxa discovered in the Fossil Hill Fauna are modeled as members of a food web. The member "shelled invertebrates" is basal to the food web and comprises primarily ammonoids but also halobiid bivalves and crustaceans. The member "nonshelled invertebrates plus fishes" ("fish") pools coleoid cephalopods such as squid and small- to medium-sized fish. See Methods and table S11 for food web members and their body masses, total biomasses, and energetic demands. (**A**) Trophic interaction matrix used for modeling energy-flux across members. Stacked bars represent the diet of predatory taxa. Filled squares within bars indicate that a taxon is taken by the predator, whereas white indicates that it is not. (**B**) Stability values calculated by the model for different combinations of total biomass of the two food web members "shelled invertebrates" and "fish." More negative stability values indicate a more stable food web. Error bars for blue dots represent model results assuming maximum and minimum body mass estimates for ichthyosaur taxa (table S11). Greater body masses result in less-stable food webs than smaller body masses. Stability values of extant food webs range between −10 and 0 (*41*). Note that we multiplied stability values by −1 for plotting. TB, total biomass (kg); red circles, the total biomass of "shelled invertebrates" equals that of "fish."

requires complementary approaches. We chose a Bayesian implementation of the Ornstein-Uhlenbeck model of trait evolution (*10*, *32*) to infer whether and when adaptive shifts of body size occurred through evolutionary time and whether smaller or larger body sizes were selected for in the new regime (Fig. 6 and fig. S12). Analyses of the selective regime also enable an evolutionary definition of giants and dwarfs. Adaptive shifts toward larger body size define clades comprising giants, with evidence for selection of larger body sizes, whereas shifts toward smaller body sizes identify dwarfs.

Analyses of the selective regime provide support for two independent shifts toward larger body sizes early in the evolution of ichthyosaurs, defining two clades of giants. One shift toward larger size is reconstructed for the branch leading to the Cymbospondylidae with *C. youngorum* sp. nov., and another shift is placed on the branch leading to the Merriamosauria (Fig. 6A). Both selective regime shifts thus occurred around the boundary between the Early and Middle Triassic. Shifts toward regimes selecting smaller body sizes were inferred for the *Hudsonelpidia brevirostris* and *Ichthyosaurus communis* tip branches (Fig. 6A).

In cetaceans, the geologically oldest shift toward a regime selecting larger body sizes occurred in the late Eocene (Fig. 6B), coinciding with the loss of functional teeth in some lineages of mysticetes (*33*). The Pliocene accelerations of body-size evolution in crown mysticetes are not detected as distinct adaptive

shifts. However, a second shift toward the evolution of large body sizes occurred in the Oligocene, fitted to the branch leading to the Pan-Physeteroidea, which evolved a raptorial feeding mode by the middle Miocene (i.e., *Brygmophyseter shigensis*). An evolutionary reduction of body sizes is inferred for the Kogiidae, a clade nested within the Pan-Physteroidea, with two living species, *Kogia breviceps* and *Kogia sima*, the pygmy and dwarf sperm whales, respectively. Another shift toward smaller body-size evolution may have occurred early in the history of cetacean evolution, in the pakicetids, yet the support for this regime shift is not nearly as strong as the other three identified shifts (Fig. 6B).

Estimates of the phylogenetic half-life [see Methods and (*10*)] suggest that ichthyosaur body size evolved an order of magnitude faster than cetacean body size when entering a new selective regime. Ichthyosaurs thus evolved large body sizes very quickly and very early on in their evolution (Fig. 4 and fig. S10), in oceans thought to be less favorable for large-bodied ecosystem members at higher trophic levels (*9*). Intriguingly, many Cymbospondylidae, which constitute one of the two shifts toward large body sizes, were present in the same assemblage, the Fossil Hill Fauna from the Fossil Hill Member of the Favret and Prida formations of northern Nevada, USA, including *C. youngorum* sp. nov.

The composition of the Fossil Hill Fauna of Nevada

The Favret Formation of the Augusta Mountains (fig. S13), Nevada, USA, spans the middle to late Anisian (Middle Triassic), a period covering more than 2 Ma (*34*). The Fossil Hill Member itself (fig. S13) is a black shale unit of variable thickness deposited in anoxic bottom waters below the storm wave base (*35*). Surface waters were well aerated and must have supported marine reptiles and abundant ammonoids and other invertebrates (*18*, *35*), but there was no benthic life, with the possible exception of halobiid bivalves. The fossils found in the unit thus represent a pelagic ecosystem and food web.

Fish fossils are only rarely preserved (*8*, *10*), but a diverse chondrichthyan fauna (*36*) from just below the Fossil Hill Member suggests that their rarity is due to preservational bias. Comprising eight taxa, the most common marine reptile fossils of the Fossil Hill Fauna are ichthyosaurs (table S10), making it the most speciose ichthyosaur fauna known. Sauropterygia are only represented by a single taxon, the pistosauroid *A. hagdorni* (*8*), in stark contrast to the rich record of sauropterygian fossils from the Tethys Middle Triassic (*37*). Among the ichthyosaurs, *C. youngorum* sp. nov. stands out because of body size, *T. saurophagis* as the oldest apex predator among secondarily aquatic amniotes (*8*), and the holotype of *C. duelferi* (*17*) as the geologically second-oldest gravid ichthyosaur. There are two more large *Cymbospondylus* species (table S3), *C. petrinus* (*38*) and *C. nichollsi* (*16*). Two species of small ichthyosaurs (*Phalarodon fraasi* and *P. callawayi*) differ in the extent of their crushing dentition (*39*). The enigmatic medium-sized *Omphalosaurus* (*40*) is interpreted as a specialized ammonoid feeder (*10*). Notably, filter feeders appear to be absent from the Fossil Hill Fauna.

The enormous size range of marine amniotes in the Fossil Hill Fauna (table S11) rivals the size range seen in modern marine mammal faunas. This range is perplexing given that Middle Triassic oceans lacked the high productivity that is thought to sustain such ecosystems today (*9*). We therefore turned to a new approach in paleontology, energy-flux modeling, to explore whether the Fossil Hill Fauna, as preserved in the fossil record, was a stable food web.

Energy-flux modeling

To test for the stability of the trophic network in the Fossil Hill Fauna, we modeled energy flux with a new tool implemented in R and derived from quantitative ecosystem ecology, "fluxweb" (*41*), which is based on allometric trophic network theory (*42*) and estimates energy fluxes in a top-down approach. In our new implementation for the fossil ecosystem, the model input is preserved taxa and census data and their estimated body masses, energy demands (table S11), and potential prey (Fig. 7A and fig. S14A) (*10*). This tool assesses the food web's stability as the smallest equilibrium total biomass of all modeled food web members by applying a predator-prey multispecies model. Small-shelled invertebrates, including ammonoids, represent the trophically basal member of the food web in the model. The small-shelled invertebrates provide energy directly or indirectly to all trophically higher food web members.

We found that the Fossil Hill food web, with its high phylogenetic diversity and morphological disparity of large to giant endothermic (*10*) ichthyosaurs, was indeed stable (Fig. 7B and fig. S16A). The preserved ammonoids alone provided sufficient energy to the food web as recorded by the fossils (fig. S16C and table S12). Sensitivity analyses demonstrate that the results of the energy-flux model are robust with respect to errors in body mass estimates for ichthyosaurs (figs. S16 and S17) and basal shelled invertebrates (fig. S15) and do not hinge on the assumption of ichthyosaurian endothermy (figs. S15 and S18). Thus, the primary production in the middle to late Anisian was sufficient to support the Fossil Hill food web, including the giant ichthyosaur *C. youngorum* sp. nov. The stability of the Fossil Hill Fauna is also consistent with the lack of pre-Triassic ichthyosaur fossils and the notion that gigantism in ichthyosaurs evolved rapidly in the first few million years of their known history because no hidden Permian history of the clade needs to be invoked (*43*).

The results from our energy-flux model challenge the hypothesis of energetic limitation of maximum body size in Mesozoic food webs (*9*) and provide insights into the functioning of a trophic network without modern primary producers. An important functional characteristic of food webs is the production rate, which is the energy that an individual stores in its body plus the energy that it allocates to reproduction, normally about one-quarter of its basal metabolic rate (*44*). Thus, the sum of the production rates of all individuals in a given trophic level defines the energy that can be transferred to the next level. Our estimates for the production rate of the preserved ammonoid population is congruent with the average production rate of multiple invertebrate species and several modern marine ecosystems (tables S12 and S13). Given that the primary producers of Mesozoic food webs were less productive than their modern equivalents (*9*) and that the trophic level of the largest ichthyosaurs equaled that of the largest living marine carnivores (*8*) and assuming that the energy losses between trophic levels had been the same as those in modern marine ecosystems, the Fossil Hill food web must have had shorter food chains than modern marine food webs.

Although the inferred production rate of all vertebrates in the Fossil Hill Fauna is similar to modern marine ecosystems (table S13), the contributions to this rate by nonamniotes and amniotes differ substantially. The production rates of modern marine amniotes (cetaceans, pinnipeds, birds) are about two magnitudes smaller than the inferred rates for the Fossil Hill Fauna (marine reptiles; table S13), whereas the rates of modern nonamniotes are up to two magnitudes larger than the modeled rates of nonamniotes from the Fossil Hill Fauna (table S13). In comparison to modern marine food webs, the trophic network of the Fossil Hill Fauna was thus dominated by marine amniotes, unlike modern marine food webs that are dominated by nonamniotes. Thus, ichthyosaurs likely occupied niches that are held by fish and whales in modern ecosystems. However, compared with their energetic demands, modeled production rates of the largest ichthyosaur species are small (table S14) and suggest that their densities were lower than suggested by the Fossil Hill Member census.

Comparison of the production rate of preserved ammonites with that modeled for the member "invertebrates" reveals an untapped

energy surplus (fig. S16C and table S12). This surplus may indicate that the food web of the Fossil Hill Fauna as now known is incomplete, perhaps lacking a bulk feeder or other taxa yet to be discovered. Alternatively, the energy surplus actually was untapped in the rather young Fossil Hill food web, being only a few million years old, and consumers of the surplus might have been added in the course of evolution later. Our energy-flux modeling indeed shows that the food web could have supported another giant marine amniote, if it fed in bulk low in the food chain (fig. S19), for example, by filter feeding (*4*). This mode of life possibly arose later in ichthyosaurs, as suggested by the existence of some Late Triassic giant toothless ichthyosaurs (*45*). Filter feeding is important in modern marine vertebrates, for example, baleen whales, the whale shark *Rhincodon typus*, and the basking shark *Cetorhinus maximus* (*46*). Even with a giant bulk feeder, the estimated energy flux from the trophic group comprising the shelled invertebrates passed on to higher trophic levels is still smaller than the estimated amount of energy provided by the preserved ammonoids (figs. S20 and S21 and table S13). Whereas there is no fossil evidence for a filter-feeding ichthyosaur in the Fossil Hill Fauna yet, the results of the energy-flux model demonstrate that this fauna was not only stable but, with its abundant ammonoids and short, amniote-dominated food chains, also set the environmental stage that led to the evolution of large body sizes early on in the evolution of ichthyosaurs.

Implications for body-size evolution of marine amniotes

The appearance of marine amniotes in the Triassic followed on the heels of amelioration of environmental conditions in the first 2 Ma of the Triassic together with a general recovery of marine ecosystems (*47*, *48*). At an estimated body mass of more than 40 tonnes and a geologic age of 246 Ma, *C. youngorum* sp. nov. was a giant marine amniote, attaining a body size comparable to that of today's ocean giants. This new giant may even have approached the size of today's largest cetacean *Balaenoptera musculus* (total length: mode of 25 m) [dataset in (*46*)], given that the Fossil Hill energy-flux model remained stable when we performed the analysis with the upper limit of the body-mass estimate of *C. youngorum* sp. nov. (24.96 m, 135,809 kg). On land, equivalent body masses did not evolve until >40 Ma later, among sauropod dinosaurs of the Jurassic. Apparently, the pelagic environment may be more conducive to the evolution of giant tetrapods. Alternatively, ecosystem recovery from the end-Permian extinction was much slower on land than in the sea. In the sea, only other ichthyosaurs in the Late Triassic and cetaceans since the late Paleogene (38 Ma ago) reached this body size again. The discovery of a giant ichthyosaur so early in the phylogenetic history of the clade underscores the existence of major selective advantages of large body size (*49*).

The Fossil Hill Fauna records a surprisingly diverse and morphologically disparate fauna of large-bodied to giant ichthyosaurs shortly after the end-Permian mass extinction. Unlike the contemporaneous faunas from the western Tethys and China, which are representative of shallow seas on continental shelves, shallow basins, and lagoons (*50*), the Fossil Hill Fauna provides a glimpse into the pelagic habitats of the Middle Triassic. We propose that ichthyosaurs initially benefited from the rapid recovery of conodonts (*51*) and pelagic ammonoid cephalopods (*47*), permitting giant body size soon after oceanic geochemical conditions had stabilized in the Middle Triassic (*48*). Ichthyosaurs may have been able to increase the total amount of resources available to them by virtue of their large eyes (*52*). Large eyes improve the range from which prey can be detected in the clear water of the pelagic realm, both in well-illuminated water near the surface and at greater depth (*53*). Proportionately large eyes seem to have evolved very early in the evolutionary history of the group (*52*), possibly enabling ichthyosaurs to better exploit their food resources. Their energetically costly endothermy could have increased their foraging capacity (speed and success) and energy intake in cold-water habitats, as found in colder geographic regions or in deeper waters (*54*). Mesothermy, which enabled body temperatures intermediate to those of ecto- and endotherms, was important in the evolution of elasmobranch gigantism (*55*).

We have identified two major evolutionary pathways to large body size in cetaceans. Size evolution in odontocetes may be linked to the evolution of raptorial feeding mode in some taxa and deep diving in others (*4*, *56*). Raptorial feeding is one of the drivers for the independent evolution of large body sizes in different clades. This is further evidenced by the earliest inferred raptorial feeder, *Ankylorhiza tiedemani*, which is considered to be the largest Oligocene odontocete with an estimated body length of 4.8 m (*57*), and in the extant delphinid *O. orca*. The evolution of echolocation, which allowed odontocetes to forage at greater depths in search of cephalopods unavailable to early cetaceans lacking a biosonar, is not directly coupled with evolution of size. Biosonar evolved more than 14 Ma before our reconstructed shift toward larger body sizes in the pan-physesteroids (*58*). In mysticetes, the initial shift toward large body sizes coincides with a loss of functional teeth in some lineages of mysticetes and presumably a switch of diet preference. In addition to evolving bulk-feeding adaptations, gigantism in mysticetes appears to be driven by global oceanic changes, particularly over the past few million years, which affected prey distribution and density (*6*, *59*). *C. youngorum* sp. nov. thus demonstrates that the lack of carbon sinks and modern primary producers is not a prerequisite for gigantism [contra (*9*)], challenging the notion that changes in ocean productivity and ecological escalation are necessary preconditions for the evolution of giant body size. Although both cetaceans and ichthyosaurs evolved very large body sizes, their respective evolutionary trajectories toward gigantism were different.

Methods

Phylogenetic analysis of ichthyosaurs

We scored the holotype specimen of *C. youngorum* sp. nov. (LACM DI 157871) into the character matrix of (*17*). This matrix (*17*) had been modified from other recent work (*7*, *60*) and appeals because the character descriptions are drawn from many different sources. We could score 40% of the 287 characters in the matrix with confidence for the new taxon (data S1) but note that many characters in this list were initially defined for post-Triassic ichthyosaurs. We edited the character-taxon matrix with Mesquite v. 3.02 (*61*). Our primary analysis (analysis I; Fig. 3 and fig. S7) included the taxon set of (*17*) with *C. youngorum* sp. nov. added (totaling 60 taxa) and was performed with TNT (*62*) using both "new technology" (search parameters: xmu=hit 20 drift 10) and "traditional" searches. See table S4 for statistics of this and the following analyses. To test for the influence of taxon sampling on the phylogenetic relationships, we used a reduced taxon set with a focus on Triassic ichthyosaurs (analysis II). This reduced taxon set is equivalent to the list of taxa used in (*63*) with the addition of *C. duelferi* and *C. youngorum* sp. nov. Note that we did not use the taxon-character matrix of (*63*), only the same taxa (data S2). A further modification was that we included three representative parvipelvians (*H. brevirostris*, *I. communis*, and *Stenopterygius quadriscissus*), instead of a clade Parvipelvia as in (*63*), to represent this derived clade (analysis III). Finally, we analyzed this matrix with PAUP* 4.1b (*64*) on a Mac computer (analysis IV). Using the heuristic search algorithm and 1000 replicates, 308 most parsimonious trees (MPTs) of 843 steps in length were retained (table S4). The strict consensus was poorly resolved, but the 50% majority rule consensus of these 308 trees shows the same topology as that found by TNT using the same matrix.

The major difference between the four analyses is in the placement of the *Cymbospondylus* clade. The analyses of the modified Klein *et al.* (*17*) taxon set (analyses I and II) recover the clade as earlier branching than, or in a trichotomy with, mixosaurids. The TNT analyses of the

modified Huang *et al.* (*63*) taxon set (analysis III) finds *Cymbospondylus* as later branching than mixosaurids and the Carnian *Toretocnemus*, *Qianichthyosaurus*, and *Californosaurus*. The PAUP analysis (analysis IV) finds the genus as later branching than the Carnian *Toretocnemus* and in an unresolved trichotomy with the Carnian *Qianichthyosaurus* and *Californosaurus*. Because the first analysis (analysis I) is the most conservative in showing the greatest similarity with recently published analyses of Triassic ichthyosaurs (*17*, *63*) and has a fully resolved strict consensus and a low number of MPTs (table S4), we consider analysis I to be our preferred hypothesis (data S3).

Total length estimate for LACM DI 157871

Of the commonly used body-size proxies—total length, skull length, humerus length, and body mass—only two, skull length and humerus length, can be directly measured in LACM DI 157871 and the other two must be estimated. We estimated the total body length of LACM DI 15787 from its humerus length after having conducted a revised regression analysis on a published dataset for Triassic ichthyosaurs (*23*) (fig. S8A and table S5). Contrary to Scheyer *et al.* (*23*), we log10-transformed total body length and humerus length before linear regression analysis, related body length to humerus length, and calculated 95% prediction intervals around the regression line (fig. S8A). This regression analysis was conducted in the software R, version 3.5.2 (*65*).

Body-mass estimates for the Fossil Hill Fauna ichthyosaurs

To estimate body mass of *C. youngorum* sp. nov. (as represented by LACM DI 15787) and the other ichthyosaur species in the Fossil Hill Fauna from total length, we generated a new dataset of body masses for Triassic ichthyosaurs (table S5) for a regression analysis of body mass on total body length (fig. S8B). This new dataset is based on a published compilation of the total body lengths of selected ichthyosaur fossil specimens (*24*). This publication further provides species-specific reference body masses for 1-m-long digital models of these fossils. To estimate body mass for each of the ichthyosaurs in the dataset (table S5), we used the respective reference mass (*24*) for each and up- or down-scaled its body mass to total body length. Note that we excluded the Jurassic ichthyosaurs from consideration because of their different body shape. After having log10-transformed total body length and body mass data from table S4, we carried out a linear regression analysis (fig. S8B) that again includes 95% prediction intervals. This regression analysis was conducted in the software R, version 3.5.2 (*65*).

Phylogenetic hypothesis for cetaceans

To look into body-size evolution in cetaceans, we compiled a comprehensive phylogeny with a special focus on early whales. This phylogeny (fig. S9) includes 250 taxa from the earliest cetaceans to representative extant species. Most large morphological or combined phylogenetic analyses of cetaceans derive from the character state matrix first published by Geisler and Sanders (*66*), with modifications made in subsequent works (*10*). However, taxon sampling in those large data matrices tends to fall short for specific groups (e.g., Ziphiidae, Pan-Physeteroidea). To have a more complete taxon sampling represented in our tree, we used the results of morphological analyses using character state matrices with less taxa, but with a more thorough sampling for specific groups, for example, for Pan-Physeteroidea (*10*). Our phylogeny is thus a combination of the results of several analyses using a variety of different morphological matrices. Although taxon sampling sometimes varies among these works, the relationships between different cetacean groups are generally stable and consistent.

The topology of our composite phylogeny agrees well with a new, comprehensive phylogenetic hypothesis presented in a recent study (*67*) regarding the position and relationships of major clades, only differing in a few minor regards. The first is that our taxon sampling is smaller, in large part because we prioritized the inclusion of species for which proxy data for body size (i.e., bizygomatic width or orbital and postorbital width) were available. Because of this requirement, we did not include *Himalayacetus subathuensis*, which is considered as the earliest cetacean, because the specimen consists of an incomplete mandible (*68*) and instead chose *P. attocki* as our earliest cetacean, which is known from more complete cranial material (*26*). The topology of Lloyd and Slater (*67*) differs from ours by having a polytomy amongst some of the more basal mysticetes (i.e., *Llanocetus denticrenatus*, *Mystacodon selenensis*). Our topology differs in the more inclusive definition of Aetiocetidae that includes *Borealodon osedax* and *Aetiocetus* (*Niparajacetus*) palmadentis, following (*69*). Concerning the odontocete section of the phylogeny, we are using a more inclusive Platanistoidea and Kentriodontidae (*70*, *71*). These differences largely derive from the analyses we used to construct our phylogeny but do not affect our results.

Time calibration of phylogenies

We used stratigraphic ranges at the stage level to time-calibrate the phylogenies (*10*). First and last appearance dates of taxa were defined as the beginning and end of each geologic stage in which the tip taxon is found (data S4 and S5). The geologic age in million years for each stage was taken from Walker *et al.* (*34*). The time calibration of the trees was performed using the R package "paleotree" (*10*) (data S6). Given that the stratigraphic occurrence of many ichthyosaurs and fossil cetaceans is not well defined, we used the most simplistic a posteriori dating approach, implemented in the "timePaleoPhy" function of "paleotree." The initial root ages were set to 251.9 and 56 Ma ago for ichthyosaurs and cetaceans, respectively. We generated sets of time-calibrated trees with the "equal" method (*10*) combined with the "minMax" option. Polytomies were resolved randomly, and for each random resolution, an internal branch length of 1 Ma was added to the tree. This is an arbitrary but, we think, reasonable estimate given the overall tree height of the clades we were working with. For plotting (Fig. 4) and applying methods that allow heterogeneity of parameters across the tree, we also generated trees with the "firstLast" setting, with terminal edges added. The "firstLast" option treats the stratigraphic bins as hard constraints. Trees shown in Fig. 4 were time calibrated with the "DatePhylo" function of the R package "strap" (*10*), also using the "equal" method, and then plotted against the geologic time scale with the "geoscalePhylo" function. The resulting time-calibrated trees for ichthyosaurs agree well with previously published phylogenies for the group, and our cetacean tree generally agrees with those of McGowen *et al.* (*72*) and Lloyd and Slater (*67*). Our divergence estimates for major clades are sometimes a few million years older (e.g., Balaenidae), younger (e.g., Balaenopteridae), or nearly identical (e.g., Pan-Physeteroidea). Such discrepancies likely reflect differences in taxon sampling and time-binning methods (*10*).

Computational analysis of body-size evolution summary

We chose skull-related metrics because skull width (bizygomatic width) is an established size proxy in cetaceans [e.g., (*73*)], as is skull length in ichthyosaurs (*7*). We used the latter dataset for ichthyosaurs, including some additional data from the recent literature and this study (data S4). Specifically, we added the skull length of two more species of *Chaohusaurus*, *C. chaoxianensis* and *C. zhangjiawanensis*, among Early Triassic taxa. Note that the specimen of *C. chaoxianensis* that provides the skull length has recently been assigned to a new species, *C. brevifemoralis*, sister taxon to *C. chaoxianensis* (data S4). Among Middle Triassic taxa, we added the specimen described in this study and the recently described *C. duelferi* from the same beds (*17*) to the dataset. The dataset of cetacean bizygomatic width relies on many sources and includes a number of new data points (data S5).

We first evaluated the fit of trait evolution models to the data for the given phylogenies

(fig. S10) with a maximum likelihood approach, using the "fitContinuous" function of the R package "geiger" (*10*) (data S6). We fit five different models with macroevolutionary relevance [(*10*) and table S7]. We ran the primary analyses over 1000 trees with slightly differing time calibration and the subsequent resampling analyses to explore the data in more depth over 100 different trees. For each iteration, we normalized branch lengths to avoid computational problems. We evaluated model fit by means of the sample-size corrected Akaike information criterion (AICc) and their corresponding Akaike weights over the entire tree sets. Akaike weights are useful because they represent the conditional probability for each of the tested models (*10*). To account for any artifactual or other nonbiological differences between our ichthyosaur and cetacean datasets, we not only performed model fitting for the entire dataset but also carried out several resampling approaches of the cetacean data. Altogether, we completed nine different model-fitting approaches (see fig. S11).

To illustrate the patterns of body-size evolution in ichthyosaurs and cetaceans, we selected a new modification of the well-established evolutionary traitgram approach (*10*). In evolutionary traitgrams, the phylogeny is combined with ancestral state reconstructions for the nodes through a projection of the tree into a space defined by the trait value on the y axis and time on the x axis (or sometimes vice versa). We intended to achieve a direct comparison of ichthyosaur and cetacean body-size evolution in the same diagram. To realize this type of visualization, we had to overcome several hurdles. First, we used different size proxies for ichthyosaurs (skull length) and cetaceans (skull width), and thus the trait values are not directly comparable in visual terms. To solve this issue, we normalized the trait values so that the smallest and largest trait values for both ichthyosaurs and cetaceans ranged from 0 to 1, respectively (Fig. 4). Second, ichthyosaurs and cetaceans originated at different times in geologic history. To facilitate their direct comparison, we started each traitgram at the root of the tree (Fig. 4), irrespective of absolute geologic time. Third, the mode of body-size evolution differs between ichthyosaurs and cetaceans, with early burst–like processes dominating ichthyosaurs and no clear pattern for the full cetacean dataset. We therefore reconstructed the traitgrams for cetaceans with the simplest evolutionary model (Brownian motion) and for ichthyosaurs using an early-burst model (fig. S10). Finally, to illustrate uncertainty stemming from the difficulties of proper time calibration, we devised and applied a new approach for traitgram plotting. The stratigraphic ranges of many extinct taxa are not well defined, and as such, the branch lengths of our trees are uncertain. To take this uncertainty into account, we plotted a range of different possible evolutionary traitgrams into the same diagram, using semitransparent colors (Fig. 4 and fig. S10). Dense, more opaque areas in the traitgram space therefore indicate evolutionary time and trait combinations with a high probability of containing the true evolutionary path of body size, whereas empty areas are less probable (Fig. 4 and fig. S10). For both ichthyosaurs and cetaceans, we also added the traitgrams that one would expect if the last appearance date were taken at face value, that is, if all taxa existed to the very end of the stage they have been documented for. These traitgrams are in white color. The evolutionary traitgram approach was implemented using the R package "phytools" (*31*) (data S6).

To identify regions of the tree with accelerated phases of evolution, we explored heterogeneity in the rate of trait evolution across the phylogeny using the function "multirateBM" of the R package "phytools" (*10*, *31*). This method is based on a penalized likelihood in which we fit a multirate Brownian motion trait evolution model in which the rate (σ^2) evolves by a correlated random-walk process. Every edge of the tree is consequently allowed to have a slightly different value of σ^2. The penalty term, which is based on the probability density of the evolutionary rates σ^2 along all branches in the tree multiplied by a user-specified coefficient denominated λ, is necessary to identify both the evolutionary rates along all the edges of the tree, as well as the rate of Brownian evolution of the rate itself. Low values for the penalty coefficient λ (e.g., 0.01) will tend to accord very little penalty to rate variation among edges, whereas higher values (e.g., λ = 100) permit the rate to differ little. Intermediate λ values (e.g., 1) balance the probability of the data under the model and the probability of the rate variation among edges, given a Brownian evolution process for rate variation. We set λ = 0.1 for both ichthyosaur and cetacean datasets and verified that the overall pattern remains consistent for $0.01 \leq \lambda \leq 1$. Given that we did not perform a full cross-validation of λ, we consider the results exploratory.

The second approach to assess body-size evolution along the edges of the phylogeny is founded on a Bayesian implementation of the Ornstein-Uhlenbeck method [R package "bayOU" (*32*)]. The Ornstein-Uhlenbeck model of trait evolution is an extension of the Brownian motion model and describes a random walk coupled with the tendency of trait values to stay in proximity of a stationary peak (θ). The tendency of the trait to remain close to the peak is measured by the parameter α, which is interpreted as the strength of selection. BayOU analyses thus characterize the selective regime of a trait.

BayOU agnostically infers whether adaptive shifts of the peak θ occurred through evolutionary time and, if so, along which edge in the phylogeny these changes likely happened. Bayesian approaches to such inferences are also less prone to error for small phylogenetic comparative datasets (N < 100) than other methods (*74*), which is useful for the ichthyosaur data. Probabilistic prior settings are summarized in table S8.

The reversible jump Markov chain Monte Carlo simulations were run for 3 million generations for the ichthyosaur data and 6 million generations for the cetacean data, of which the first 30% was discarded as burn-in. We ensured that independent chains had converged on similar regions in the parameter space by Gelman's R for log likelihood, σ^2, and α (fig. S12 and table S9). We also checked for convergence by plotting the posterior probabilities for shifts along branches against each other. If convergence was achieved, the posterior probabilities should fall along a line with a slope of 1 (fig. S12). We considered an adaptive shift well-supported if its respective posterior probability was far outside the main distribution of posterior probabilities for all branches.

Among ichthyosaurs, four branches with mean posterior probabilities of 0.41 to 0.86 [39 to 81 times greater than their prior probability (0.01)] were chosen (see table S8). The mean estimate of α is 0.4, indicating a phylogenetic half-life [$\ln(2)/\alpha$] of 1.73 Ma. Applied to cymbospondylids, the ichthyosaurs in this clade evolved halfway from their previous adaptive peak of 135 mm to their new peak (1443 mm) in 1.73 Ma, which is congruent with the fossil record. The phylogenetic half-life of cetacean skull width is 17.3 Ma. Among cetaceans, three branches with mean posterior probabilities of 0.84 to 0.95 [413 to 472 times greater than their prior probability (0.002)] were chosen (see table S8). Another branch received a posterior probability of 0.36, far less than the other three shifts and not as clearly separated from the distribution of posterior probabilities. If supported, this adaptive shift would suggest that the Pakicetidae shifted toward smaller body size, with a new peak 62-mm skull width as opposed to the ancestral state estimate of 121 mm. Other interesting patterns are apparent in cetacean size evolution, such as the missing middle size classes in the Oligocene and repeated increases of body-size ranges in different clades, but fully exploring them is beyond the scope of this study.

Energy-flux modeling summary

To test whether the composition of the pelagic Fossil Hill Fauna as found in the fossil record (fig. S13 and table S10) represents a functional and stable food web (hypothesis 1, standard scenario), we modeled energy flux using a

new tool from quantitative ecosystem ecology implemented in R, "fluxweb" (*41*), which is based on allometric trophic network theory (*42*). Importantly, it derives energy fluxes in a top-down approach. The "fluxweb" tool calculates the food web's stability (*41*) by applying a predator-prey multispecies model and searching for equilibrium total biomasses of food web members. More negative stability values indicate a more stable food web, that is, the smallest equilibrium total biomass across all food web members is larger than zero, whereas a positive stability value indicates that at least one food web member is extinct under equilibrium.

Model input is census data of preserved taxa (all of which are pelagic) as members of the food web, estimated body masses, energetic demands (table S11), and potential prey (Fig. 7A and fig. S14). For modeling, we added two further members that are lowest in the food web and pool different kinds of animal taxa. We pooled taxa because we lack sufficient information on the parameter values required for modeling each taxon individually, and we aimed to keep our model as simple as possible. The first additional member is "shelled invertebrates" ("invertebrates" for short). This member is basal in the modeled food web and comprises primarily ammonoids but also halobiid bivalves and crustaceans. It pools the trophically lowest invertebrates of the Fossil Hill Fauna preserved in the fossil record. The "invertebrates" directly or indirectly provide energy to all other members of the food web. How this energy is produced in a basal member is ignored in "fluxweb" (*41*), and thus modeling of predator-prey relations within basal food web members is not possible in this tool. The second additional member is "pooled nonshelled invertebrates and fish" ("fish" for short). This member pools coleoid cephalopods such as squid and small- to medium-sized fish. It thus comprises the preserved and unpreserved taxa of the lowest trophic level on which the majority of all other trophically higher members fed (Fig. 7A and fig. S14). For the member "fish," we implemented potential within-group predator-prey relations by allowing that "fish" feed on "fish" (Fig. 7A and fig. S14) because larger fish could potentially have fed on smaller fish and squid. We further assume in our model that energy demands of all ichthyosaur taxa equaled that of modern endothermic vertebrates (*10*), whereas demands of all other taxa either equaled that of ectothermic vertebrates or of ectothermic invertebrates.

For the two members "invertebrates" and "fish," we evaluated our model for 15 different combinations of total biomasses because we lack reliable information on their total biomasses from the fossil record. Because the body masses estimated for the ichthyosaur taxa used in our model and total biomasses derived from these all have large margins of error and because the ichthyosaurs are the largest animals of the Fossil Hill Fauna, we evaluated each of the 15 combinations for standard body mass, lower limit of body mass, and upper limit of body mass based on 95% prediction intervals (table S11). A sensitivity analysis of body mass of basal "invertebrates," for which the fossil record documents a body mass variation of about five orders of magnitude, showed that potential errors had an extremely small impact on stability values and energy fluxes calculated by our model (see below and fig. S15). In the three combinations, in which the total biomass of "invertebrates" equaled that of "fish," the average energy loss between two trophic levels always turned out to be the highest across the 15 combinations, and estimated losses of around 40% were clearly unrealistic from a modern perspective (figs. S16B and S19, B and F). We thus restricted our ecological and evolutionary interpretation of modeling results to the other 12 combinations in which the total biomass of the "invertebrates" was larger than that of the "fish."

Although we assumed that energy demands of all ichthyosaurs in the Fossil Hill Fauna conformed to modern endotherms [table S11 and (*10*)], we nevertheless wanted to test model sensitivity to this assumption. We thus reran the model for all combinations assuming ectothermic and mesothermic (*54*) ichthyosaurs. Endothermy in any of the taxa in the model (table S11) results in a higher energy consumption than ectothermy and mesothermy (fig. S15). We considered mesothermy as a metabolic strategy for ichthyosaurs because several modern marine macropredators show this strategy—for example, tunas, swordfish, and lamnid sharks (*54*)—and because mesothermy was important in the evolution of elasmobranch gigantism (*55*).

The "fluxweb" tool provides different allometric equations on mass-specific metabolic rates for implementing different energetic demands (i.e., physiological losses; fig. S15). The ectothermic and the endothermic vertebrate metabolic types of "fluxweb" only differ in their use of scaling normalization constants: 19.5 for the endothermic metabolic type versus 18.18 for the ectothermic metabolic type (table S11). To implement mesothermy, we used the arithmetic mean of the two, 18.84. The scaling exponent of all three metabolic types is −0.29 (table S11). To assess the effect of the ectothermic and mesothermic metabolic types on stability values and energy fluxes, we reran the model for mesothermic and ectothermic ichthyosaurs while keeping everything else as in the standard scenario (fig. S18).

To narrow down which total biomass of the basal "invertebrates" are the most realistic for the food web of the Fossil Hill Fauna, we estimated the total biomass of ammonoids found in the Fossil Hill Member of the Augusta Mountains from field data (table S12). We are aware of the limitations of this approach because ammonoids usually accumulate in certain layers, are not evenly distributed throughout the Fossil Hill Member, and comprise different species. To estimate total ammonoid biomass, we conducted a census by shell diameter on one randomly chosen surface of 1 m^2. Recognizing five size classes, we found the following abundances: I, >23 cm diameter, one individual; II, 23 to 6 cm, 28 individuals; III, 6 to 3 cm, 49 individuals; IV, 3 to 1 cm, 49 individuals; and V, <0.5 cm, 28 individuals. To estimate the biomass of these individuals, we used an extant *Nautilus belauensis* with a 23-cm shell diameter and a mass of 1.675 kg (*75*) as a proxy and linearly downscaled its size and mass to estimate the weight of the smaller ammonoid individuals. This calculation and census of a total of 155 individuals yielded an average body mass of about 10 g. This is the rounded mean calculated from frequencies of individuals in size classes and body masses corresponding to the respective lower-size class boundaries (except for class V, for which we used a 0.5-cm shell diameter). The abundances and body masses of the five size classes further yielded about 2.7 kg as an estimate of ammonoid biomass per square meter. A field census of random samples in the Augusta Mountains suggested that 15% (0.540 km^2), 25% (0.900 km^2), or 30% (1.080 km^2) of the Fossil Hill outcrop area (3.6 km^2) is covered by ammonoids, which corresponds to a total biomass in this area of 1458 kg, 2430 kg, or 2916 kg, respectively. We used the estimated average ammonoid body mass (10 g) for the member "invertebrates" in the "fluxweb" model (table S11).

To assess a potential impact of the body mass assumed for "invertebrates," we evaluated the model with a body mass of "invertebrates" of 0.02 g (size class V) and of 1.675 kg (size class I) for the standard scenario. Our rationale was that this member provides the energy to all other trophically higher members and that their body-mass range documented in the fossil record covers nearly five orders of magnitude. However, for both body masses, the stability values and energy fluxes obtained were nearly identical to those obtained for the standard mass of 10 g. With respect to this observation, it is important that the body mass assumed for any modeled food web member (i) is used to calculate its physiological losses (model parameter X_i). Except for "invertebrates" and "fish," body masses of all members were also used to calculate their total biomasses as a product of body mass and number of counted individuals (table S11). Because the total biomass of "invertebrates" and "fish" is fixed in each model run, only physiological losses are

altered by changing body masses. An evaluation of the equations used for physiological losses shows that whereas the body-mass range covered by the food web members spans about 10 orders of magnitude (the mass corresponding to the smallest ammonite size class V is 0.02 g, and the upper body mass limit of *C. youngorum* sp. nov. is 135.81 tonnes), the variation in losses is smaller than one magnitude over this huge range (fig. S15). For the members "fish," coelacanth fish, and *A. hagdorni*, we estimated body masses from related and similar taxa because we lack any information about their body mass and its variation in the Fossil Hill Fauna. We therefore refrained from carrying out a sensitivity analysis for body mass for these three members. Based on the arguments given above, modeling results must also be very insensitive to the body mass assumed for the members "fish" and coelacanth fish. There is only one coelacanth fish individual in the Fossil Hill Fauna, and thus its total biomass equals its body mass (table S11). However, the census for *A. hagdorni* is three individuals, and any change in the body mass assumed for this member results in a change in its total biomass. Thus, we cannot exclude the possibility that this change will alter stability values and energy fluxes. We anticipate that these values will be in the range found for ichthyosaur members with a similar body mass and total biomass.

Modeling the Fossil Hill Fauna with a hypothetical giant bulk feeder

Averaged across the 12 combinations of total biomasses of "invertebrates" and "fish," the production rate of "invertebrates" was a magnitude smaller than that of the preserved ammonoids (fig. S16C and table S12). We thus conducted a second modeling experiment (hypothesis 2) in which we added a hypothetical blue whale–sized endothermic ichthyosaur (200 tonnes, about 30 m long) to the Fossil Hill food web model while keeping everything else as in the standard scenario. This super giant bulk feeder exclusively fed on either "invertebrates" (fig. S14B) or "fish" (fig. S14C). For both feeding strategies, the hypothesized food web was again stable (fig. S19A), and for each chosen combination of total biomass of "invertebrates" and "fish," food web stability was somewhat higher than in the standard scenario (fig. S20A). When averaged across the possible combinations of total biomasses of "invertebrates" and "fish," the supergiant that fed exclusively on "invertebrates" consumed an increase in outgoing energy from "invertebrates" of 2.6×10^{-7} kJ/year and the supergiant that fed exclusively on "fish" consumed 1.7×10^{7} kJ/year (fig. S21). A detailed description of our model, a justification of model assumptions, and output variables inferred is in the supplementary materials (*10*).

REFERENCES AND NOTES

1. G. J. Vermeij, Gigantism and its implications for the history of life. *PLOS ONE* **11**, e0146092 (2016). doi: 10.1371/journal.pone.0146092; pmid: 26771527
2. N. P. Kelley, N. D. Pyenson, Vertebrate evolution. Evolutionary innovation and ecology in marine tetrapods from the Triassic to the Anthropocene. *Science* **348**, aaa3716 (2015). doi: 10.1126/science.aaa3716; pmid: 25883362
3. A. H. Knoll, M. J. Follows, A bottom-up perspective on ecosystem change in Mesozoic oceans. *Proc. Biol. Sci.* **283**, 20161755 (2016). doi: 10.1098/rspb.2016.1755; pmid: 27798303
4. J. A. Goldbogen *et al*., Why whales are big but not bigger: Physiological drivers and ecological limits in the age of ocean giants. *Science* **366**, 1367–1372 (2019). doi: 10.1126/science.aax9044; pmid: 31831666
5. W. Gearty, C. R. McClain, J. L. Payne, Energetic tradeoffs control the size distribution of aquatic mammals. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 4194–4199 (2018). doi: 10.1073/pnas.1712629115; pmid: 29581289
6. G. J. Slater, J. A. Goldbogen, N. D. Pyenson, Independent evolution of baleen whale gigantism linked to Plio-Pleistocene ocean dynamics. *Proc. Biol. Sci.* **284**, 20170546 (2017). doi: 10.1098/rspb.2017.0546; pmid: 28539520
7. B. C. Moon, T. L. Stubbs, Early high rates and disparity in the evolution of ichthyosaurs. *Commun. Biol.* **3**, 68 (2020). doi: 10.1038/s42003-020-0779-6; pmid: 32054967
8. N. B. Fröbisch, J. Fröbisch, P. M. Sander, L. Schmitz, O. Rieppel, Macropredatory ichthyosaur from the Middle Triassic and the origin of modern trophic networks. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 1393–1397 (2013). doi: 10.1073/pnas.1216750110; pmid: 23297200
9. P. M. Hull, Emergence of modern marine ecosystems. *Curr. Biol.* **27**, R466–R469 (2017). doi: 10.1016/j.cub.2017.04.041; pmid: 28586680
10. See supplementary materials.
11. C. v. Linnaeus, *Systema Naturae. Regnum Animale* (L. Salvius, Stockholm, ed. 10, 1758).
12. H. F. Osborn, The reptilian subclasses Diapsida and Synapsida and the early history of the Diaptosauria. *Mem. Am. Mus. Nat. Hist.* **1**, 449–507 (1903).
13. H. M. D. de Blainville, Systéme de'Herpetologie. *Nouv. Ann. Mus. Hist. Nat. Paris* **4**, 233–296 (1835).
14. J. Leidy, Notice of some reptilian remains from Nevada. *Proc. Acad. Nat. Sci. Philadelphia* **20**, 177–178 (1868).
15. P. M. Sander, The large ichthyosaur *Cymbospondylus buchseri*, sp. nov., from the Middle Triassic of Monte San Giorgio (Switzerland), with a survey of the genus in Europe. *J. Vertebr. Paleontol.* **9**, 163–173 (1989). doi: 10.1080/02724634.1989.10011750
16. N. Fröbisch, P. M. Sander, O. Rieppel, A new species of *Cymbospondylus* (Diapsida, Ichthyosauria) from the Middle Triassic of Nevada and re-evaluation of the skull osteology of the genus. *Zool. J. Linn. Soc.* **147**, 515–538 (2006). doi: 10.1111/j.1096-3642.2006.00225.x
17. N. Klein, L. Schmitz, T. Wintrich, P. M. Sander, A new cymbospondylid ichthyosaur (Ichthyosauria) from the Middle Triassic (Anisian) of the Augusta Mountains, Nevada, USA. *J. Syst. Palaeontology* **18**, 1167–1191 (2020). doi: 10.1080/14772019.2020.1748132
18. C. Monnet, H. Bucher, New middle and late Anisian (Middle Triassic) ammonoid faunas from Northwestern Nevada (USA): Taxonomy and biochronology. *Foss. Strat.* **52**, 1–121 (2005).
19. B. Moon, A new phylogeny of ichthyosaurs (Reptilia: Diapsida). *J. Syst. Palaeontology* **17**, 129–155 (2019). doi: 10.1080/14772019.2017.1394922
20. J. A. Massare, Tooth morphology and prey preference of Mesozoic marine reptiles. *J. Vertebr. Paleontol.* **7**, 121–137 (1987). doi: 10.1080/02724634.1987.10011647
21. R. Motani, Evolution of fish-shaped reptiles (Reptilia: Ichthyopterygia) in their physical environments and constraints. *Annu. Rev. Earth Planet. Sci.* **33**, 12.11–12.26 (2005).
22. C. McGowan, Giant ichthyosaurs of the Early Jurassic. *Can. J. Earth Sci.* **33**, 1011–1021 (1996). doi: 10.1139/e96-077
23. T. M. Scheyer, C. Romano, J. Jenks, H. Bucher, Early Triassic marine biotic recovery: The predators' perspective. *PLOS ONE* **9**, e88987 (2014). doi: 10.1371/journal.pone.0088987; pmid: 24647136
24. S. Gutarra *et al*., Effects of body plan evolution on the hydrodynamic drag and energy requirements of swimming in ichthyosaurs. *Proc. Biol. Sci.* **286**, 20182786 (2019). doi: 10.1098/rspb.2018.2786; pmid: 30836867
25. R. Motani *et al*., A basal ichthyosauriform with a short snout from the Lower Triassic of China. *Nature* **517**, 485–488 (2015). doi: 10.1038/nature13866; pmid: 25383536
26. S. Nummela, T. Hussain, J. G. M. Thewissen, Cranial anatomy of Pakicetidae (Cetacea, Mammalia). *J. Vertebr. Paleontol.* **26**, 746–759 (2006). doi: 10.1671/0272-4634(2006)26[746:CAOPCM]2.0.CO;2
27. J. G. M. Thewissen, J. D. Sensor, M. T. Clementz, S. Bajpai, Evolution of dental wear and diet during the origin of whales. *Paleobiology* **37**, 655–669 (2011). doi: 10.1666/10038.1
28. M. A. MacIver, L. Schmitz, U. Mugan, T. D. Murphey, C. D. Mobley, Massive increase in visual range preceded the origin of terrestrial vertebrates. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E2375–E2384 (2017). doi: 10.1073/pnas.1615563114; pmid: 28270619
29. J. M. Fahlke, K. A. Bastl, G. M. Semprebon, P. D. Gingerich, Paleoecology of archaeocete whales throughout the Eocene: Dietary adaptations revealed by microwear analysis. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **386**, 690–701 (2013). doi: 10.1016/j.palaeo.2013.06.032
30. L. N. Cooper *et al*., Aquatic habits of cetacean ancestors: Integrating bone microanatomy and stable isotopes. *Integr. Comp. Biol.* **56**, 1370–1384 (2016). doi: 10.1093/icb/icw119; pmid: 27697778
31. L. J. Revell, Phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* **3**, 217–223 (2012). doi: 10.1111/j.2041-210X.2011.00169.x
32. J. C. Uyeda, L. J. Harmon, A novel Bayesian method for inferring and interpreting the dynamics of adaptive landscapes from phylogenetic comparative data. *Syst. Biol.* **63**, 902–918 (2014). doi: 10.1093/sysbio/syu057; pmid: 25077513
33. C. M. Peredo, N. D. Pyenson, C. D. Marshall, M. D. Uhen, Tooth loss precedes the origin of baleen in whales. *Curr. Biol.* **28**, 3992–4000.e2 (2018). doi: 10.1016/j.cub.2018.10.047; pmid: 30503622
34. J. D. Walker, J. W. Geissman, S. A. Bowring, L. E. Babcock, "Geologic time scale v. 5.0" (Geological Society of America, 2018).
35. K. M. Nichols, N. J. Silberling, Stratigraphy and depositional history of the Star Peak Group (Triassic), northwestern Nevada. *Spec. Pap. Geol. Soc. Am.* **178**, 1–73 (1977).
36. G. Cuny, O. Rieppel, P. M. Sander, The shark fauna from the Middle Triassic (Anisian) of North-Western Nevada. *Zool. J. Linn. Soc.* **133**, 285–301 (2001). doi: 10.1111/j.1096-3642.2001.tb00627.x
37. O. Rieppel, *Handbook of Paleoherpetology / Sauropterygia I.: Placodontia, Pachypleurosauria, Nothosauroidea, Pistosauroidea* (Friedrich Pfeil, 2000).
38. J. C. Merriam, *Triassic Ichthyosauria, with Special Reference to the American Forms* (Memoirs of the University of California, The University Press, 1908).
39. L. Schmitz, P. M. Sander, G. W. Storrs, O. Rieppel, New Mixosauridae (Ichthyosauria) from the Middle Triassic of the Augusta Mountains (Nevada, USA) and their implications for mixosaur taxonomy. *Palaentographica A* **270**, 133–162 (2004).
40. J. C. Merriam, Preliminary note on a new marine reptile from the Middle Triassic of Nevada. *Univ. Calif. Publication Bull. Dept. Geol.* **5**, 5–79 (1906).
41. B. Gauzens *et al*., Fluxweb: An R package to easily estimate energy fluxes in food webs. *Methods Ecol. Evol.* **10**, 270–279 (2018). doi: 10.1111/2041-210X.13109
42. N. D. Martinez, Allometric trophic networks from individuals to socio-ecosystems: Consumer-resource theory of the ecological elephant in the room. *Front. Ecol. Evol.* **8**, 92 (2020). doi: 10.3389/fevo.2020.00092
43. R. Motani, D.-Y. Jiang, A. Tintori, C. Ji, J.-D. Huang, Pre- versus post-mass extinction divergence of Mesozoic marine reptiles dictated by time-scale dependence of evolutionary rates. *Proc. Biol. Sci.* **284**, 20170241 (2017). doi: 10.1098/rspb.2017.0241; pmid: 28515201
44. J. H. Brown, J. F. Gillooly, A. P. Allen, V. M. Savage, G. B. West, Toward a metabolic theory of ecology. *Ecology* **85**, 1771–1789 (2004). doi: 10.1890/03-9000
45. E. L. Nicholls, M. Manabe, Giant ichthyosaurs of the Triassic – A new species of *Shonisaurus* from the Pardonet Formation (Norian, Late Triassic) of British Columbia. *J. Vertebr. Paleontol.* **24**, 838–849 (2004). doi: 10.1671/0272-4634(2004)024[0838:GIOTTN]2.0.CO;2
46. C. R. McClain *et al*., Sizing ocean giants: Patterns of intraspecific size variation in marine megafauna. *PeerJ* **3**, e715 (2015). doi: 10.7717/peerj.715; pmid: 25649000
47. A. Brayard *et al*., Unexpected Early Triassic marine ecosystem and the rise of the Modern evolutionary fauna. *Sci. Adv.* **3**,

e1602159 (2017). doi: 10.1126/sciadv.1602159; pmid: 28246643
48. H. Song, P. B. Wignall, A. M. Dunhill, Decoupled taxonomic and ecological recoveries from the Permo-Triassic extinction. *Sci. Adv.* **4**, eaat5091 (2018). doi: 10.1126/sciadv.aat5091; pmid: 30324133
49. P. M. Sander *et al*., Biology of the sauropod dinosaurs: The evolution of gigantism. *Biol. Rev. Camb. Philos. Soc.* **86**, 117–155 (2011). doi: 10.1111/j.1469-185X.2010.00137.x; pmid: 21251189
50. M. J. Benton *et al*., Exceptional vertebrate biotas from the Triassic of China, and the expansion of marine ecosystems after the Permo-Triassic mass extinction. *Earth Sci. Rev.* **137**, 85–128 (2014). doi: 10.1016/j.earscirev.2014.08.004
51. M. J. Orchard, Conodont diversity and evolution through the latest Permian and Early Triassic upheavals. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **252**, 93–117 (2007). doi: 10.1016/j.palaeo.2006.11.037
52. R. Motani, B. M. Rothschild, W. J. Wahl Jr., Large eyeballs in diving ichthyosaurs. *Nature* **402**, 747 (1999). doi: 10.1038/45435
53. D.-E. Nilsson, E. Warrant, S. Johnsen, Computational visual ecology in the pelagic realm. *Philos. Trans. R. Soc. London Ser. B* **369**, 20130038 (2014). doi: 10.1098/rstb.2013.0038; pmid: 24395965
54. J. M. Grady *et al*., Metabolic asymmetry and the global diversity of marine predators. *Science* **363**, eaat4220 (2019). doi: 10.1126/science.aat4220; pmid: 30679341
55. C. Pimiento, J. L. Cantalapiedra, K. Shimada, D. J. Field, J. B. Smaers, Evolutionary pathways toward gigantism in sharks and rays. *Evolution* **73**, 588–599 (2019). doi: 10.1111/evo.13680; pmid: 30675721
56. A. T. Boersma, N. D. Pyenson, *Albicetus oxymycterus*, a new generic name and redescription of a basal physeteroid (Mammalia, Cetacea) from the Miocene of California, and the evolution of body size in sperm whales. *PLOS ONE* **10**, e0135551 (2015). doi: 10.1371/journal.pone.0135551; pmid: 26651027
57. R. W. Boessenecker, M. Churchill, E. A. Buchholtz, B. L. Beatty, J. H. Geisler, Convergent evolution of swimming adaptations in modern whales revealed by a large macrophagous dolphin from the Oligocene of South Carolina. *Curr. Biol.* **30**, 3267–3273.e2 (2020). doi: 10.1016/j.cub.2020.06.012; pmid: 32649912
58. J. H. Geisler, M. W. Colbert, J. L. Carew, A new fossil species supports an early origin for toothed whale echolocation. *Nature* **508**, 383–386 (2014). doi: 10.1038/nature13086; pmid: 24670659
59. J. A. Goldbogen, P. T. Madsen, The evolution of foraging capacity and gigantism in cetaceans. *J. Exp. Biol.* **221**, jeb166033 (2018). doi: 10.1242/jeb.166033; pmid: 29895582
60. E. E. Maxwell, D. Y. Cortés, P. Patarroyo, M. L. P. Ruge, A new specimen of *Platypterygius sachicarum* (Reptilia, Ichthyosauria) from the Early Cretaceous of Colombia and its phylogenetic implications. *J. Vertebr. Paleontol.* **39**, e1577875 (2019). doi: 10.1080/02724634.2019.1577875
61. W. P. Maddison, D. R. Maddison, Mesquite: A modular system for evolutionary analysis. Version 3.02 (2015); www.mesquiteproject.org.
62. P. A. Goloboff, J. S. Farris, K. C. Nixon, TNT, a free program for phylogenetic analysis. *Cladistics* **24**, 774–786 (2008). doi: 10.1111/j.1096-0031.2008.00217.x
63. J.-D. Huang *et al*., The new ichthyosauriform *Chaohusaurus brevifemoralis* (Reptilia, Ichthyosauromorpha) from Majiashan, Chaohu, Anhui Province, China. *PeerJ* **7**, e7561 (2019). doi: 10.7717/peerj.7561; pmid: 31565558
64. D. Swofford, *PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods). Version 4.0b10* (Sinauer Associates, 2002).
65. R Core Team, R software, version 3.5.2 (R Foundation for Statistical Computing, Vienna, Austria, 2020); www.R-project.org/.
66. J. H. Geisler, A. E. Sanders, Morphological evidence for the phylogeny of Cetacea. *J. Mamm. Evol.* **10**, 23–129 (2003). doi: 10.1023/A:1025552007291
67. G. T. Lloyd, G. J. Slater, A total-group phylogenetic metatree for Cetacea and the importance of fossil data in diversification analyses. *Syst. Biol.* **70**, 922–939 (2021). doi: 10.1093/sysbio/syab002; pmid: 33507304
68. S. Bajpai, P. D. Gingerich, A new Eocene archaeocete (Mammalia, Cetacea) from India and the time of origin of whales. *Proc. Natl. Acad. Sci. U.S.A.* **95**, 15464–15468 (1998). doi: 10.1073/pnas.95.26.15464; pmid: 9860991
69. A. E. Hernández Cisneros, J. Velez-Juarbe, Paleobiogeography of the North Pacific toothed mysticetes (Cetacea: Aetiocetidae): A key on the Oligocene cetacean distributional patterns. *Palaeontology* **64**, 51–61 (2021). doi: 10.1111/pala.12507
70. M. Viglino, M. R. Buono, R. E. Fordyce, J. I. Cuitiño, E. M. G. Fitzgerald, Anatomy and phylogeny of the large shark-toothed dolphin *Phoberodon arctirostris* Cabrera, 1926 (Cetacea: Odontoceti) from the early Miocene of Patagonia (Argentina). *Zool. J. Linn. Soc.* **185**, 511–542 (2019). doi: 10.1093/zoolinnean/zly053
71. C. M. Peredo, M. D. Uhen, M. D. Nelson, A new kentriodontid (Cetacea: Odontoceti) from the early Miocene Astoria Formation and a revision of the stem delphinidan family Kentriodontidae. *J. Vertebr. Paleontol.* **38**, e1411357 (2018). doi: 10.1080/02724634.2017.1411357
72. M. R. McGowen *et al*., Phylogenomic resolution of the cetacean tree of life using target sequence capture. *Syst. Biol.* **69**, 479–501 (2020). doi: 10.1093/sysbio/syz068; pmid: 31633766
73. N. D. Pyenson, S. N. Sponberg, Reconstructing body size in extinct crown Cetacea (Neoceti) using allometry, phylogenetic methods and tests from the fossil record. *J. Mamm. Evol.* **18**, 269–288 (2011). doi: 10.1007/s10914-011-9170-1
74. N. Cooper, G. H. Thomas, C. Venditti, A. Meade, R. P. Freckleton, A cautionary note on the use of Ornstein Uhlenbeck models in macroevolutionary studies. *Biol. J. Linn. Soc. Lond.* **118**, 64–77 (2016). doi: 10.1111/bij.12701; pmid: 27478249
75. D. L. Pisor, *Registry of World Record Size Shells* (ConchBooks, ed. 5, 2008).

ACKNOWLEDGMENTS

We thank D. Goodreau and E. Durazo [all from Dinosaur Institute (DI), Natural History Museum of Los Angeles County (LACM)] for preparation of the specimen; M. Walsh (DI) for curation; and S. Abramowicz (DI) and G. Oleschinski (Bonn) for photography. S. Abramowicz is acknowledged for drafting the summary figure. E. Maxwell (Stuttgart) kindly made the data matrix from (*60*) available before final publication and provided helpful technical advice. The specimen was collected under Bureau of Land Management (BLM) Paleontological Resources Use Permit N-92625. The help of BLM staff at the Winnemucca field office is gratefully acknowledged. We thank the reviewers for their constructive comments. T. and B. Young and their Great Basin Brewery (Reno, NV, USA) generously supported our fieldwork by contributions in kind and monetary contributions. **Funding:** This research was funded in part by grants from the Deutsche Forschungsgemeinschaft (project numbers 388659338 and 264173172) to P.M.S. and by the National Geographic Society (grant number 9599-14) and the W. M. Keck Science Department to L.S. Great Basin Brewery (Reno, NV) supported our fieldwork and the preparation of the specimen at the LACM. Additional support was provided by the DI under the direction of L. Chiappe. The work of L.J.R. on phytools is supported by the National Science Foundation (DEB-1350474 and DBI-1759940). **Author contributions:** The overall idea, concept, and approach were developed by P.M.S., L.S., and T.W. Morphological descriptions and phylogenetic analysis were performed by P.M.S., N.K., T.W., and L.S. Evolutionary rate analysis was performed by L.S., J.V.J., and L.J.R. Energy-flux modeling was performed by E.M.G., N.K., and P.M.S. The manuscript was written by P.M.S. and L.S. with contributions from E.M.G., J.V.J., L.J.R., N.K., and T.W. All authors read, edited, and discussed the manuscript and participated in data collection. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The holotype specimen of *C. youngorum* sp. nov. is owned by the US Department of the Interior and housed in the collections of the Dinosaur Institute, Natural History Museum of Los Angeles County (Los Angeles, CA, USA) as LACM DI 157871. All data analyzed in this study are available in the main text and supplementary materials. R scripts are available at data S6 and S8.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abf5787
Supplementary Text
Figs. S1 to S21
Tables S1 to S14
References (*76*–*178*)
MDAR Reproducibility Checklist
Data S1 to S8

6 November 2020; accepted 26 October 2021
10.1126/science.abf5787

RESEARCH ARTICLES

MOLECULAR BIOLOGY

Pol IV and RDR2: A two-RNA-polymerase machine that produces double-stranded RNA

Kun Huang[1,2]†, Xiao-Xian Wu[1]†, Cheng-Li Fang[1,2]†, Zhou-Geng Xu[2,3]†, Hong-Wei Zhang[1,2], Jian Gao[2,3], Chuan-Miao Zhou[3], Lin-Lin You[1,2], Zhan-Xi Gu[1,2], Wen-Hui Mu[4], Yu Feng[5]*, Jia-Wei Wang[3]*, Yu Zhang[1]*

DNA methylation affects gene expression and maintains genome integrity. The DNA-dependent RNA polymerase IV (Pol IV), together with the RNA-dependent RNA polymerase RDR2, produces double-stranded small interfering RNA precursors essential for establishing and maintaining DNA methylation in plants. We determined the cryo–electron microscopy structures of the Pol IV–RDR2 holoenzyme and the backtracked transcription elongation complex. These structures reveal that Pol IV and RDR2 form a complex with their active sites connected by an interpolymerase channel, through which the Pol IV–generated transcript is handed over to the RDR2 active site after being backtracked, where it is used as the template for double-stranded RNA (dsRNA) synthesis. Our results describe a 'backtracking-triggered RNA channeling' mechanism underlying dsRNA synthesis and also shed light on the evolutionary trajectory of eukaryotic RNA polymerases.

To date, five multiple-subunit DNA-dependent RNA polymerases (msDdRPs) have been identified in eukaryotes. Polymerase (Pol) I, Pol II, and Pol III are conserved in all eukaryotic organisms (*1*), whereas Pol IV and Pol V are specific to land plants (*2*–*5*). Decades of studies have demonstrated that Pol IV and Pol V—together with an RNA-dependent RNA polymerase (RDR2)—are key members of the plant RNA-directed DNA methylation (RdDM) pathway. Pol IV, in association with RDR2, generates double-stranded RNAs (dsRNAs) at the RdDM locus of genomic DNA (*6*–*8*). Subsequently, the dsRNA precursors are processed into 24-nucleotide (nt) small interfering RNAs (siRNAs) that pair with Pol V–generated scaffold RNAs to recruit the DNA methyltransferase DRM2 for DNA methylation (*9*–*13*). There is increasing evidence to suggest that the RdDM pathway is essential for repressing transposable elements, establishing genome imprinting, and maintaining genome integrity (*13*–*15*).

Although it has been proposed that Pol IV is derived from Pol II (*16*, *17*), the catalytic activity (*18*–*20*), regulatory transcription factors (*21*–*25*), and targeting of genomic loci of Pol IV differ radically from those of Pol II (*22*, *23*, *26*, *27*). Precisely how Pol IV transcription initiates, elongates, and terminates, and how Pol IV partners with RDR2 to produce dsRNA precursors at specific genomic loci, remain elusive. To address these questions, we solved cryo–electron microscopy (cryo-EM) structures of the *Arabidopsis* Pol IV–RDR2 holoenzyme and Pol IV–RDR2 backtracked transcription elongation complex (bTEC).

A cryo-EM structure of the *Arabidopsis* Pol IV–RDR2 holoenzyme

Direct purification of low-abundance, multiple-subunit protein complexes for structural study has not yet been achieved in plants. To circumvent this problem, we purified endogenous Pol IV from an engineered *Arabidopsis thaliana* cell line (T87) stably expressing 3xFLAG-tagged NRPD1, the largest subunit of Pol IV (fig. S1, A to C) (*28*). Liquid chromatography–mass spectrometry (LC-MS/MS) and SDS-polyacrylamide gel electrophoresis (PAGE) analysis of the purified complex shows the presence of Pol IV subunits and RDR2, suggesting that we obtained a stable Pol IV–RDR2 complex (fig. S1D and table S1), in agreement with previous reports that RDR2 is associated with Pol IV in vivo (*18*, *25*). The purified Pol IV–RDR2 complex is catalytically active in extending a Pol IV RNA primer with a bipartite DNA scaffold (fig. S1E) (*18*).

The structure of the Pol IV–RDR2 holoenzyme was determined by using the single-particle cryo-EM method at a nominal resolution of 3.86 Å (fig. S2 and movie S1). Pol IV resembles a crab claw, the characteristic shape for all msDdRPs (Fig. 1) (*29*). Structure superimposition of Pol IV with other msDdRPs confirmed a closer relationship between Pol IV and Pol II than that between Pol I and Pol III (Fig. 2B and fig. S4, A to C) (*16*, *17*). The structure of RDR2 is superimposable on the crystal structure of the catalytic domain of a cellular-encoded RNA-dependent RNA polymerase (cRdRP) from *Neurospora crassa* [root mean square deviation (RMSD) 1.3 Å for all Cα atoms; fig. S4D] (*30*). Pol IV and RDR2 interact with each other through a large otherwise solvent-exposed surface of 2265 Å^2, where the secondary channel of Pol IV approaches the active center of RDR2 (Fig. 1C and movie S2).

Pol IV has a characteristic active center cleft

Pol IV resembles Pol II in overall shape but differs in the conformations of key motifs in its active center cleft (Fig. 2, A to C, and fig. S4, E to G) (*31*). No large domain rearrangement of the 10 subunits of Pol IV occurs compared with Pol II, with the exception of the zinc ribbon domain of Pol IV NRPB9 subunit (NRPB9 zinc ribbon), which relocates to a new position ~40 Å from where the RPB9 zinc ribbon is located in Pol II and establishes extensive interactions with NRPD1 funnel helices (NRPD1 FH) at the new site (Fig. 2F). The Pol IV–specific DeCL-like domain (Fig. 1A) is disordered in our structure. Similar to Pol II, Pol IV contains a main cleft for accommodating DNA and RNA at the deep bottom of the two lobes (Fig. 2A). Protein sequence alignment reveals that several key motifs inside the main cleft of Pol IV—the rudder loop, the lid loop, the bridge helix, and the trigger loop—exhibit consensus sequences distinct from those in Pol II (fig. S6 and table S3) (*31*). In our Pol IV–RDR2 structure, these key structural elements either become disordered or adopt distinct conformations compared with those in the Pol II structure (Fig. 2C and fig. S4, E to G, and I).

In the main cleft of Pol IV, the rudder and lid loops exhibit disorder, rendering the entire active center cleft wide open and indicating that the two loops are flexible and exist as an ensemble of multiple conformations (Fig. 2C and fig. S9B). In the active site of Pol IV, the bridge helix bends in the middle (Fig. 2C and fig. S9C). The two-residue deletion at the central region compared with Pol II accounts for the bent bridge helix of Pol IV, and also likely hinders a bent-to-straight conformational transition (necessary for the nucleotide addition cycle of Pol II) (fig. S6). The trigger loop of Pol IV adopts an open conformation in our structure (Fig. 2C and fig. S9D). It bears a six-residue deletion and sequence variation at a key position for catalysis; i.e., Pol IV NRPD1 contains S898 at the position of the conserved histidine (H1085 in *Saccharomyces cerevisiae* Pol II–RPB1 subunit) in Pol II (fig. S6) (*32*). In summary, our structures reveal substantial conformational differences in key structural motifs at the Pol IV active site (i.e., the rudder

[1]Key Laboratory of Synthetic Biology, CAS Center for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology and Ecology, Chinese Academy of Sciences, Shanghai 200032, China. [2]University of Chinese Academy of Sciences, Beijing 100049, China. [3]National Key Laboratory of Plant Molecular Genetics, CAS Center for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology and Ecology, Chinese Academy of Sciences, Shanghai 200032, China. [4]Key Laboratory of Plant Stress Biology, State Key Laboratory of Cotton Biology, School of Life Sciences, Henan University, Kaifeng 475004, China. [5]Department of Biophysics, and Department of Pathology of Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Hangzhou 310058, China.
***Corresponding authors. Email: yufengjay@zju.edu.cn (Y.F.); jwwang@sippe.ac.cn (J.-W.W.); yzhang@sippe.ac.cn (Y.Z.)**
†These authors contributed equally to this work.

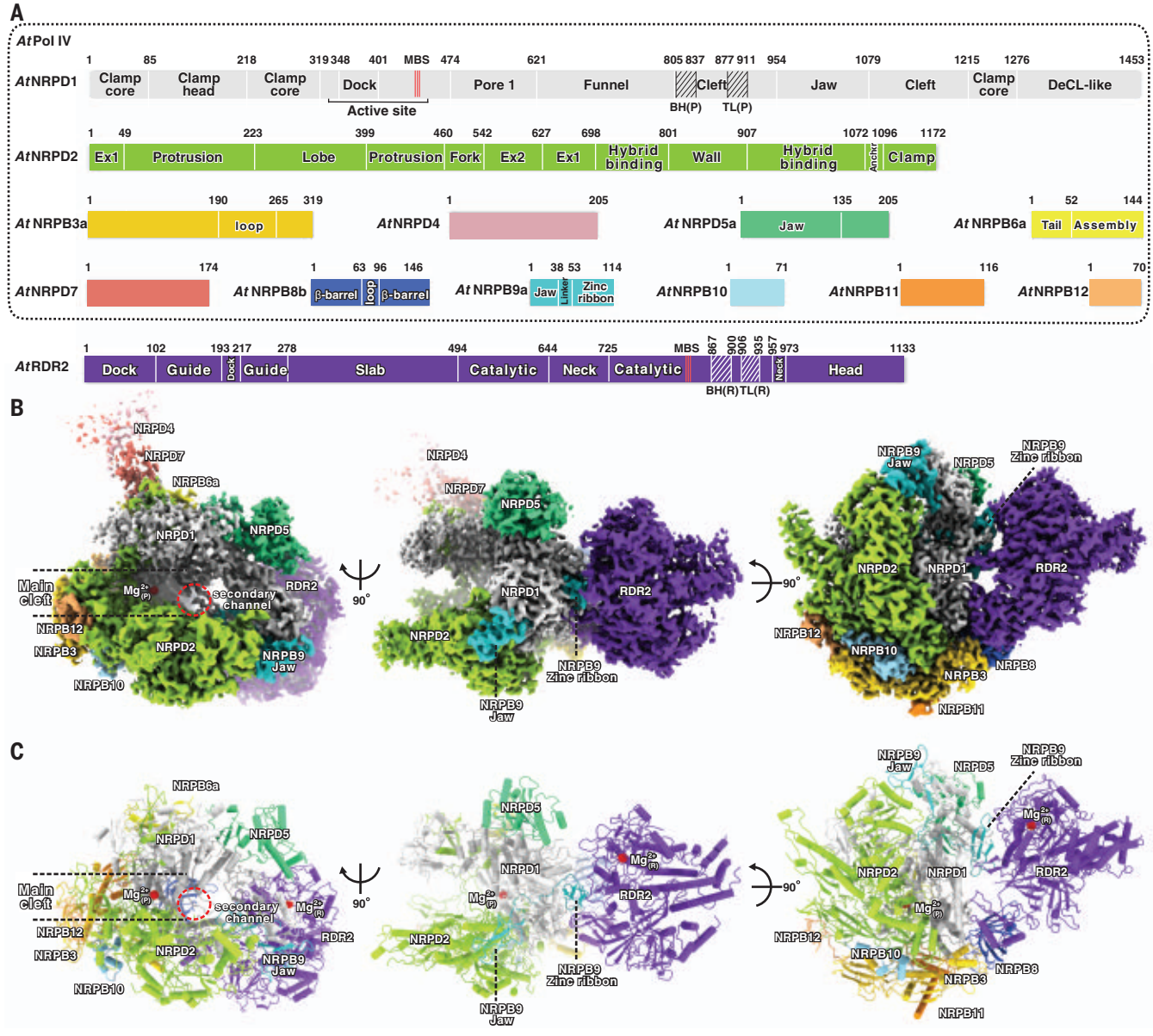


Fig. 1. The cryo-EM structure of the *A. thaliana* Pol IV–RDR2 holoenzyme. (**A**) A schematic presentation of Pol IV–RDR2 subunits. MBS, Mg^{2+} binding site; BH(P), Pol IV bridge helix; TL(P), Pol IV trigger loop; BH(R), RDR2 bridge helix; TL(R), RDR2 trigger loop. (**B**) Cryo-EM map of the Pol IV–RDR2 holoenzyme. (**C**) Structure model of the Pol IV–RDR2 holoenzyme. Subunits are colored as in (A). Red dashed circle shows the Pol IV secondary channel. Black dashed lines indicate the main cleft of pol IV. Red dots indicate catalytic Mg^{2+} in the Pol IV ($Mg^{2+}_{(P)}$) and RDR2 ($Mg^{2+}_{(R)}$) active centers.

loop, the lid loop, the bridge helix, and the trigger loop) compared with those of Pol II.

Pol IV lacks surfaces for recruiting Pol II transcription factors

Aside from the key motifs inside the active center cleft, Pol IV shows greater sequence and conformational differences on its surface compared with Pol II, as it lacks anchoring surfaces for Pol II initiation and elongation factors. For example, TFIIB interacts with conserved residues of multiple surface patches on Pol II (fig. S7). By contrast, all of the surface patches used for recruiting TFIIB are either absent or with nonconsensus residues in Pol IV (Fig. 2D and fig. S7). These structural features explain previous observations that TFIIB, TFIIF, and other Pol II transcription initiation factors were not detected in the LC-MS/MS analysis of the immunoprecipitated Pol IV (*16*). Because TFIIB is the central hub for docking of TBP/TFIID-bound promoter DNA to Pol II (*33*, *34*), the loss of TFIIB-anchoring surfaces on Pol IV supports the idea that Pol IV does not recognize Pol II promoters and uses a mechanism for transcription initiation distinct from that of Pol II (*21*–*24*).

The Pol IV–RDR2 holoenzyme structure also suggests that Pol IV is defective in binding to TFIIS, the proofreading elongation factor of Pol II that plays a key role in maintaining transcription fidelity and resolving backtracking (*35*). TFIIS is an extended protein with two domains (domain II and III) connected by an interdomain linker. Domain II docks on the Pol II–RPB1 jaw domain, and the fully extended linker helps domain III insert into the active site through the secondary channel of Pol II (fig. S4L) (*36*, *37*). Pol IV does not have the anchor helix on the NRPD1 jaw domain for anchoring domain II of TFIIS (Fig. 2E) (*38*). The entry route of the TFIIS interdomain linker is occupied by the relocated NRPB9 zinc ribbon of Pol IV (Fig. 2E and fig. S4, K and L). The sequence and structure features of Pol IV

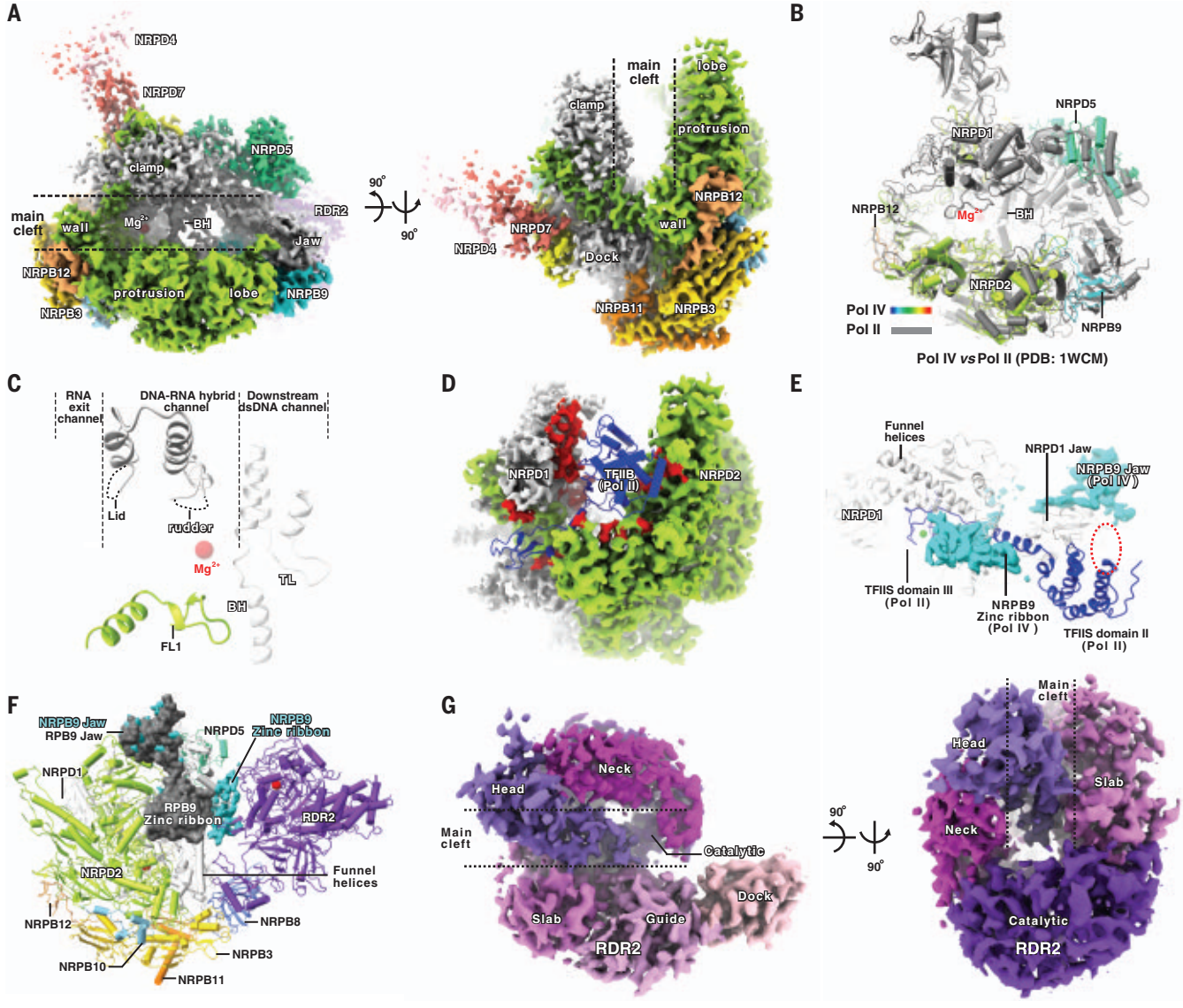


Fig. 2. Structural features of Pol IV and RDR2. (**A**) Cryo-EM map of Pol IV. Colors as in Fig. 1. (**B**) Structure superimposition between *A. thaliana* Pol IV [(colored as in (A)] and yeast Pol II (gray; PDB: 1WCM). (**C**) The lid, rudder, and fork loop 1 (FL1), bridge helix (BH), and trigger loop (TL) of Pol IV. (**D**) Pol IV mutates contact surfaces for TFIIB. The red surface patches highlight mutated residues of Pol IV on the corresponding TFIIB-contact surface of Pol II. The Yeast TFIIB (blue) is superimposed onto the structure of Pol IV. (**E**) TFIIS is sterically incompatible with Pol IV NRPB9 (cyan). The Pol II–bound TFIIS (PDB: 1Y1V; blue) is superimposed onto the structure of Pol IV–RDR2 , and a dashed oval indicates the position of the anchor helix of Pol II. (**F**) NRPB9 Zinc ribbon (cyan) is relocated in Pol IV compared with that (gray) in Pol II (PDB: 1WCM). (**G**) Cryo-EM map of RDR2.

thus imply an incompatible interaction between TFIIS and Pol IV, consistent with the absence of TFIIS in previous and our LC-MS/MS analysis of affinity-purified Pol IV. Our results suggest that TFIIS is unable to cleave backtracked RNA of Pol IV and resolve arrested Pol IV as it does during Pol II transcription (*16*).

RDR2 rides on the saddle-like surface of Pol IV

RDR2 belongs to a large family of cRdRPs and plays a central role in the RNA interference (RNAi) pathway in plants, fungi, and metazoans (*30*, *39*). Like those of msDdRPs, the active center cleft of RDR2 is located at the bottom of two lobes (Fig. 2G). The slab and guide domains comprise one lobe, and the neck domain comprises the other. The catalytic center of RDR2 is composed of structural elements similar to the D loop, the bridge helix, and the trigger loop of msDdRPs (fig. S4H). The structural similarities imply that RDR2 likely accommodates nucleic acids in the active site and catalyzes the RNA extension in a manner similar to that of msDdRP.

RDR2 rides on the saddle-like surface of Pol IV created by the NRPB3 loop domain, the NRPB8 β-barrel domain, the NRPB9 zinc ribbon, the NRPD1 cleft, and the NRPD1 FH (Fig. 3, A and B, and movie S3). RDR2 wedges its dock domain into the Pol IV saddle (Fig. 3B). The conserved residue W51 of the W loop of the RDR2 dock domain is secured by a hydrophobic pocket on NRPD1 FH; the nearby residues R48 and K52 of the W loop of the RDR2 dock domain make potential salt-bridge interactions with residue D710 of NRPD1 FH (Fig. 3C and fig. S5A). The guide and slab domains of RDR2 also insert two loops (the R loop and the YF loop) into the Pol IV saddle. Residue R207 of the R loop likely makes salt-bridge bonds with E641 and E642 of NRPD1 FH (Fig. 3C and fig. S5B); residues Y441 and F442 of the YF loop insert into hydrophobic pockets on the interface of the NRPD1 FH and NRPB9 zinc ribbon (Fig. 3, C and D, fig. S5C, and movie S3).

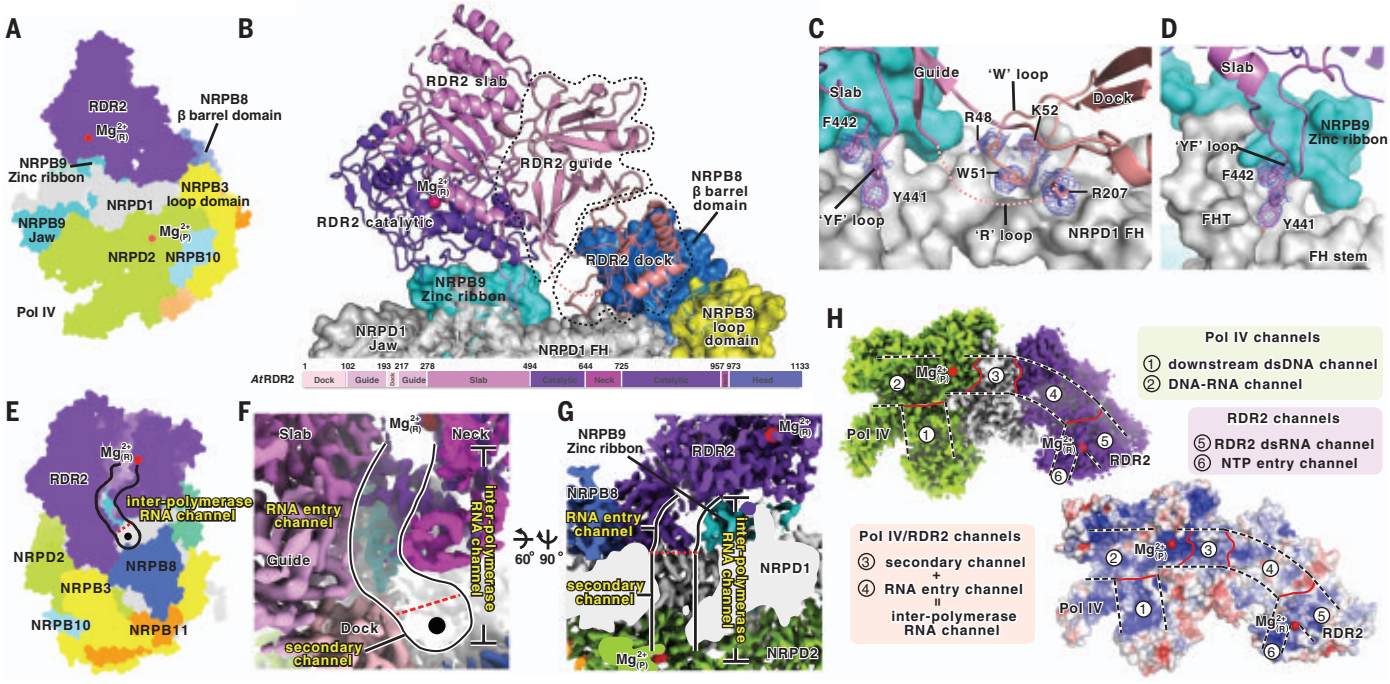


Fig. 3. RDR2 rides on the saddle-like surface of Pol IV. (**A**) Sketch of the Pol IV–RDR2 interaction. (**B**) The interface of RDR2 and Pol IV. Pol IV is shown as cartoon and half-transparent surface; RDR2 is shown as cartoon. FH, funnel helices. (**C** and **D**) Detailed interaction between RDR2 (cartoon) and Pol IV (surface). Blue mesh, cryo-EM map. (**E**) Sketch of the interpolymerase RNA channel in the Pol IV–RDR2 complex. The circle with a center dot shows the secondary channel of Pol IV perpendicular to the paper plane. The black lines delineate the path of the RDR2 RNA entry channel. The red dashed line shows the boundaries of the RDR2 RNA entry channel and the Pol IV secondary channel. (**F**) A scaled-up view of the interpolymerase RNA channel. (**G**) An intersection showing the Pol IV secondary channel. (**H**) Cryo-EM map (left) and electrostatic surfaces (right) (blue, positively charged; red, negatively charged) of Pol IV-RDR2 (certain domains omitted for clarity) showing nucleic acid channels. Channels are delineated by black dotted lines, and channel boundaries are indicated by red lines. Red dots, catalytic Mg^{2+} of Pol IV ($Mg^{2+}{}_{(P)}$) and of RDR2 ($Mg^{2+}{}_{(R)}$).

Additionally, the catalytic domain of RDR2 makes extensive contact with the relocated NRPB9 zinc ribbon (Fig. 3B and movie S3).

Protein sequence alignment analysis suggests that the interface residues on NRPD1 FH are conserved in Pol IV and distinct from those of Pol II, conferring the specificity of RDR2 to Pol IV (fig. S6) (*18*). The reciprocal interface residues on RDR2 are conserved in RDR2 homologs but not in RDR2 paralogs (i.e., RDR1 and RDR6 in *A. thaliana*), thus supporting a characteristic role of RDR2 in the Pol IV–RdDM pathway (fig. S8) (*13*, *18*).

An interpolymerase RNA channel bridges the active sites of Pol IV and RDR2

The Pol IV–RDR2 structure shows that the active sites of the two RNAPs are directly connected. The secondary channel of Pol IV is extended by the dock and guide domains of RDR2 that form an RNA entry channel leading to the active site of RDR2 (Fig. 3, E to G). The secondary channel of Pol IV and the RNA entry channel of RDR2 thereby form a continuous half-closed interpolymerase RNA channel connecting the active site of Pol IV, where Pol IV–generated RNA (Pol IV RNA) is synthesized, and the active site of RDR2 where the RDR2-generated RNA (RDR2 RNA) is synthesized using Pol IV RNA as the template (Fig. 3H). The secondary channel of msDdRPs primarily serves as the entrance for substrate NTP into the active site but also accommodates RNA at the backtracked state (a state in which msDdRPs move backward on the DNA template), which is induced under certain circumstances such as NTP misincorporation, transcription pausing, and DNA damage (*40*).

The presence of an interpolymerase RNA channel in the Pol IV–RDR2 complex implicates a possible backtracking-triggered RNA channeling model for dsRNA synthesis by Pol IV–RDR2, in which the single-stranded Pol IV RNA is directly delivered to RDR2 through the interpolymerase RNA channel triggered by Pol IV backtracking. The 3′ terminus of the Pol IV RNA strand is displaced from the Pol IV active site, further threaded through the interpolymerase RNA channel to the active site of RDR2, and then directly used as the template for synthesis of the RDR2 RNA strand of dsRNA. This method of RNA delivery is energetically favorable as the backtracked Pol IV RNA gains more electrostatic attraction interactions en route to the RDR2 active site through the interpolymerase RNA channel (Fig. 3H). Such a mechanism predicts that initiation of RDR2 RNA synthesis does not require previous termination and release of single-stranded Pol IV RNA from the Pol IV RNA exit channel, thereby preventing Pol IV RNA from degradation by multiple cellular RNA nucleases, conferring the substrate specificity of RDR2 by directly receiving the Pol IV RNA template from inside the complex, and ensuring that dsRNA production is restricted to RdDM loci of genomic DNA. Three experimental results support the backtracking-triggered RNA channeling model: First, Pol IV–RDR2 produces RDR2 RNA much more efficiently on a backtracking-prone 3′-tailed double-stranded DNA template than on a single-stranded DNA template (fig. S11A) (*36*, *41*). Second, the Pol IV–RDR2 complex protects a much longer region of backtracked RNA than Pol II (fig. S11B). Third, Pol IV–RDR2 extends the RDR2–RNA primer only when backtracked Pol IV RNA is ≥23 nt (fig. S11C). The backtracking-triggered RNA channeling model has also been proposed in a recent study (*42*).

Backtracked Pol IV RNA is threaded into the RDR2 active site for synthesis of RDR2 RNA

To further demonstrate the backtracking-triggered RNA channeling model, we reconstituted a Pol IV–RDR2 bTEC with a synthetic nucleic acid scaffold comprising four strands: (i) The nontemplate DNA strand comprises a

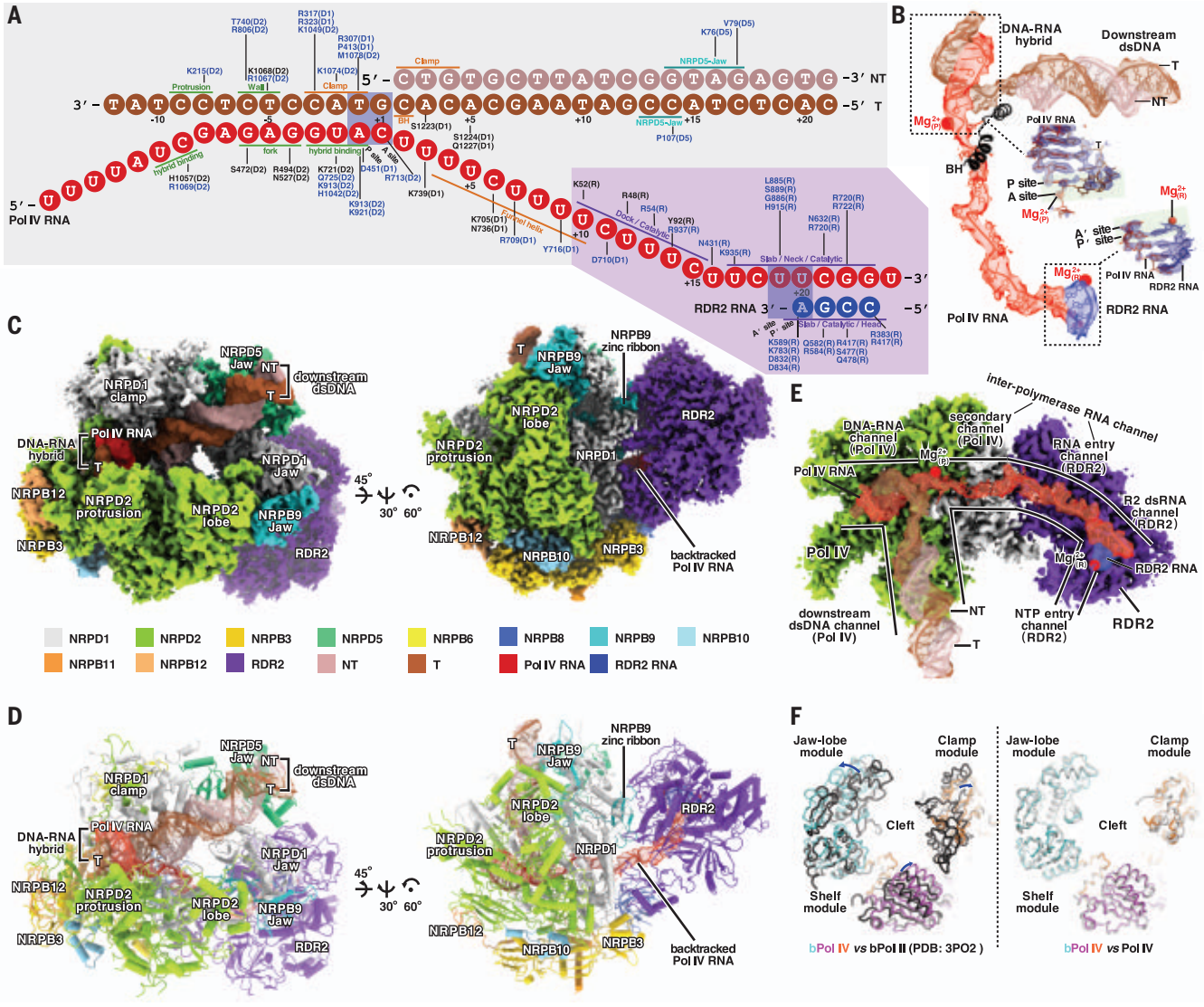


Fig. 4. The cryo-EM structure of Pol IV–RDR2 bTEC. (**A**) The nuclei acid scaffold for structure determination. Pol IV–RDR2 residues are linked to their contact nucleotides. (Conserved residues between Pol IV and Pol II and conserved residues of RDR2 in various plant species are colored in blue; nonconserved residues are colored in black; D1, D2, and R in the parenthesis represent NRPD1, NRPD2, and RDR2, respectively). 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**) The unsharpened cryo-EM map of the nucleic acid scaffold. The blue mesh shows the sharpened map for nucleotides in the active centers of Pol IV and RDR2. (**C**) The cryo-EM map of Pol IV–RDR2 bTEC in top and side view orientations. (**D**) The structure model in cartoon presentation of Pol IV–RDR2 bTEC. A transparent map of the nucleic acid scaffold is also shown. (**E**) Cryo-EM map for the nucleic acid scaffold in Pol IV–RDR2 channels. (**F**) (Left) Superimposition between Pol IV–RDR2 bTEC (bPol IV, colored as above) and Pol II backtracked TEC (bPol II; black); (right) superimposition between Pol IV–RDR2 bTEC (bPol IV, colored as above) and Pol IV–RDR2 holoenzyme (gray).

20-nt duplex-forming segment; (ii) the template DNA strand comprises a 20-nt duplex-forming segment followed by a 13-nt tail; (iii) the Pol IV RNA strand comprises a 9-nt segment complimentary to the template DNA strand as well as a 7-nt 5′ tail and a 23-nt 3′ tail; (iv) the RDR2 RNA strand comprises a 4-nt primer complimentary to the Pol IV RNA strand (Fig. 4A). The reconstituted Pol IV–RDR2 bTEC is active in extending the 4-nt RDR2–RNA primer (fig. S1F). The structure of the complex was determined at a nominal resolution of 3.16 Å (fig. S3). The cryo-EM map reveals strong, sharp signals for the 9-base pair (bp) RNA-DNA hybrid (–9 to –1) at the active center cleft of Pol IV, and a 4-bp dsRNA (+20 to +23) in the active center cleft of RDR2, allowing assignment of the nucleotide register in Pol IV–RDR2 channels. The cryo-EM map also reveals a less strong but unambiguous signal for dsDNA in the downstream dsDNA channel of Pol IV and for single-stranded RNA (ssRNA) in the interpolymerase RNA channel of Pol IV-RDR2. No signal was observed for Pol IV RNA in the Pol IV–RNA exit channel (Fig. 4, B to E, and movie S4).

The conformational states of key structural elements, as well as the interactions of Pol IV and RDR2, are identical between the structures of Pol IV–RDR2 bTEC and Pol IV–RDR2 holoenzyme (0.86 RMSD for all Cα atoms). Engagement of the nucleic acid scaffold does not induce clamp closure of Pol IV as seen in DNA-bound Pol II (Fig. 4F). The bridge helix of Pol IV remains in the kinked conformation, as seen in the structure of Pol IV–RDR2 holoenzyme (Fig. 5A; fig. S4, E, F, and J; and fig. S9C). The rudder and lid loops remain disordered, resulting in fully exposed DNA

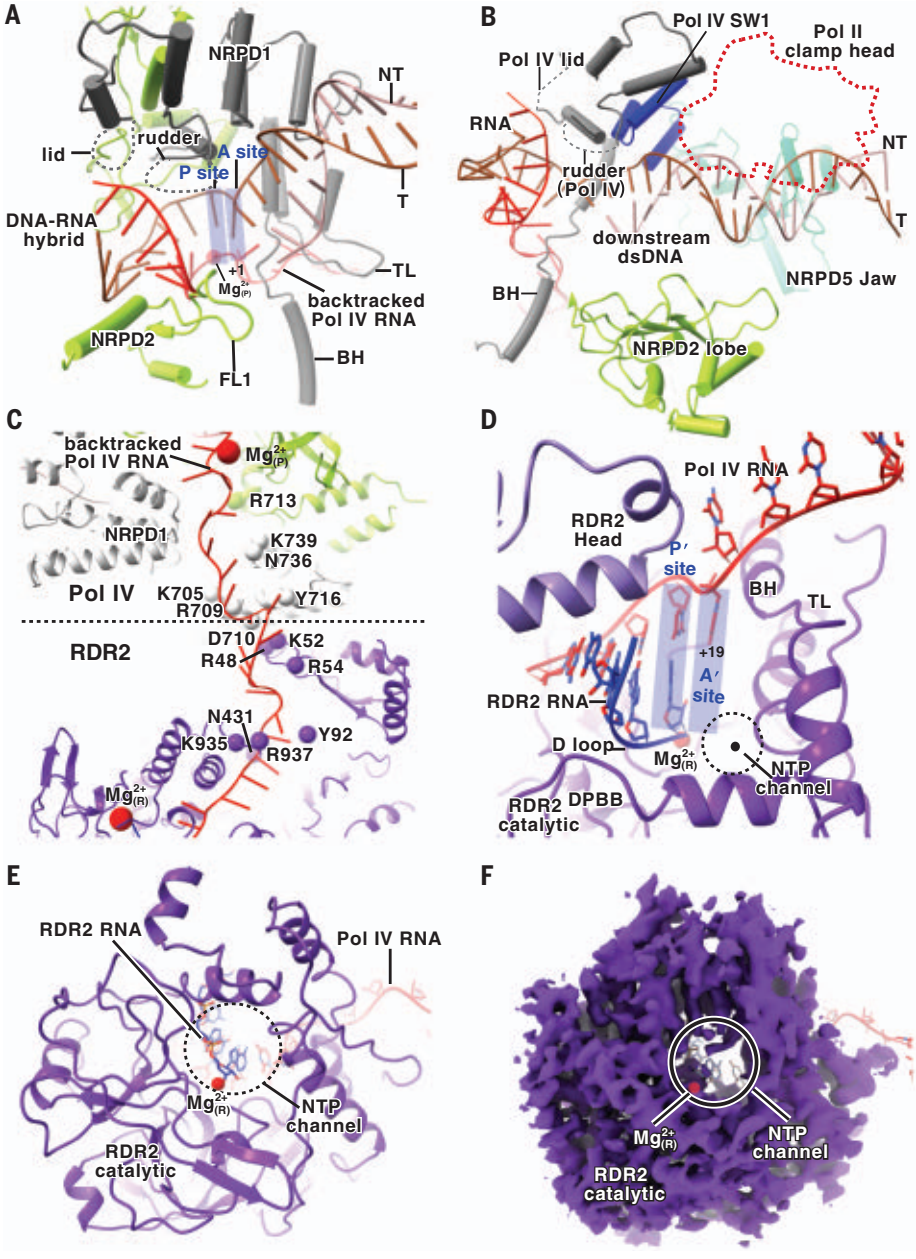


Fig. 5. Detailed interaction between Pol IV-RDR2 and the nucleic acid scaffold. (**A**) The RNA-DNA hybrid in the Pol IV active center cleft. Blue rectangles highlight the product site (P site) and nucleotide insertion site (A site) of the Pol IV active center. (**B**) Interaction between Pol IV and downstream dsDNA. Red dashes indicate the position of Pol II clamp-head domain. Gray dashes represent disordered the lid and rudder loops. (**C**) The backtracked Pol IV RNA makes interaction with polar residues (shown as purple and gray dots) in the interpolymerase RNA channel. The dashed line indicates the interface of Pol IV and RDR2. (**D**) RDR2 active center. Blue rectangles highlight the product site (P′ site) and nucleotide insertion site (A′ site) of the RDR2 active center. DPBB, conserved double-psi β-barrel domain of msRNAPs. (**E**) Illustration and (**F**) cryo-EM map of RDR2 showing the NTP channel.

and RNA in the active center cleft (Fig. 5A and figs. S4F and S9B). By contrast, the rudder and lid loops in structures of the normal and backtracked Pol II elongation complexes exhibit confined conformation that secures and maintains the transcription bubble, and thereby confers high processivity to Pol II (fig. S4G) (*43*). Pol IV interacts loosely with dsDNA in the downstream dsDNA channel (Fig. 5B, fig. S9E, and fig. S10, D and E), likely due in part to a lack of dsDNA grippers at the channel sidewalls (the rudder, clamp head, and lobe loops) used by Pol II to grip the phosphate backbones of dsDNA (figs. S6 and S9E) (*44*). In summary, our structures reveal that key structural elements in the active center cleft adopt conformations distinct from those of Pol II, and that the distinct conformations remain unchanged upon engagement of the backtracked nucleic acid scaffold.

The Pol IV RNA (nucleotides of +1 to +24) is threaded through the interpolymerase RNA channel from the Pol IV active site into the RDR2 active site (Fig. 4, B and E, and movie S4). The channel proximally accommodates 17-nt ssRNA (+2 to +18). The polar residues (K705, R709, D710, Y716, N736, and K739 of Pol IV NRPD1; R713 of Pol IV NRPD2; and R48, K52, R54, Y92, N431, K935, and R937 of RDR2) in the interpolymerase RNA channel likely delineate the path of backtracked RNA by allowing potential H-bond and salt-bridge interactions with its phosphate backbone (Figs. 4A and 5C, and fig. S10, F and G). RDR2 accommodates the 4-bp A-form dsRNA at its active center cleft in a post-translocation state (Fig. 4, A and B, Fig. 5D, fig. S10H, and movie S5). The 3′-OH group of the nucleotide A+20 of RDR2 RNA makes a coordinate bond with the catalytic Mg^{2+} constrained by the catalytic D-triad of RDR2 (D830, D832, and D834; Fig. 5D). The template nucleotide at the A′ site of RDR2 (i.e., U+19 of Pol IV RNA) is unpaired, awaiting its cognate NTP that likely diffuses into the active site through a shallow NTP channel (Fig. 4A and Fig. 5, D to F). The RDR2 head domain stabilizes short dsRNA from the top of the dsRNA groove but blocks the path of nascent RNA, suggesting that it must be displaced upon further RNA extension (Fig. 5D). We propose that the head module is functionally analogous to domain $\sigma_{3.2}$ of bacteria-initiating σ factor or the finger domain of eukaryotic initiation factor TFIIB, which initially preorganize the template strand for efficient initiation of RNA synthesis but are displaced after RNA extends to 4 to 7 nt forming a stable RNA-DNA hybrid (*45*, *46*).

Disrupting the Pol IV–RDR2 interface causes defects in DNA methylation

On the basis of our structures, we propose that the Pol IV–RDR2 interaction is a prerequisite for synthesis of dsRNA precursors (*20*). To validate this hypothesis, we determined whether disruption of Pol IV–RDR2 interactions would affect RdDM readouts. We generated transgene lines in which the FLAG-tagged, wild-type NRPD1, the NRPD1 (ΔFHT) mutant-bearing deletion of the nonconserved tip of NRPD1 FH, and the NRPD1 (ΔFHT; M5)-bearing deletion of the nonconserved tip of NRPD1 FH and the alanine substitution of five residues at the Pol IV–RDR2 interface (fig. S5D) were expressed from its native promoter in the *nrpd1* mutant. The wild-type, *nprd1* mutant, and resultant two transgenic lines were then subjected to bisulfite sequencing and small RNA sequencing. P_{NRPD1}*::NRPD1* fully rescued defects in the RdDM of the *nrpd1* mutant; P_{NRPD1}*::NRPD1* (ΔFHT) slightly reduced the 24-nt siRNA production and DNA

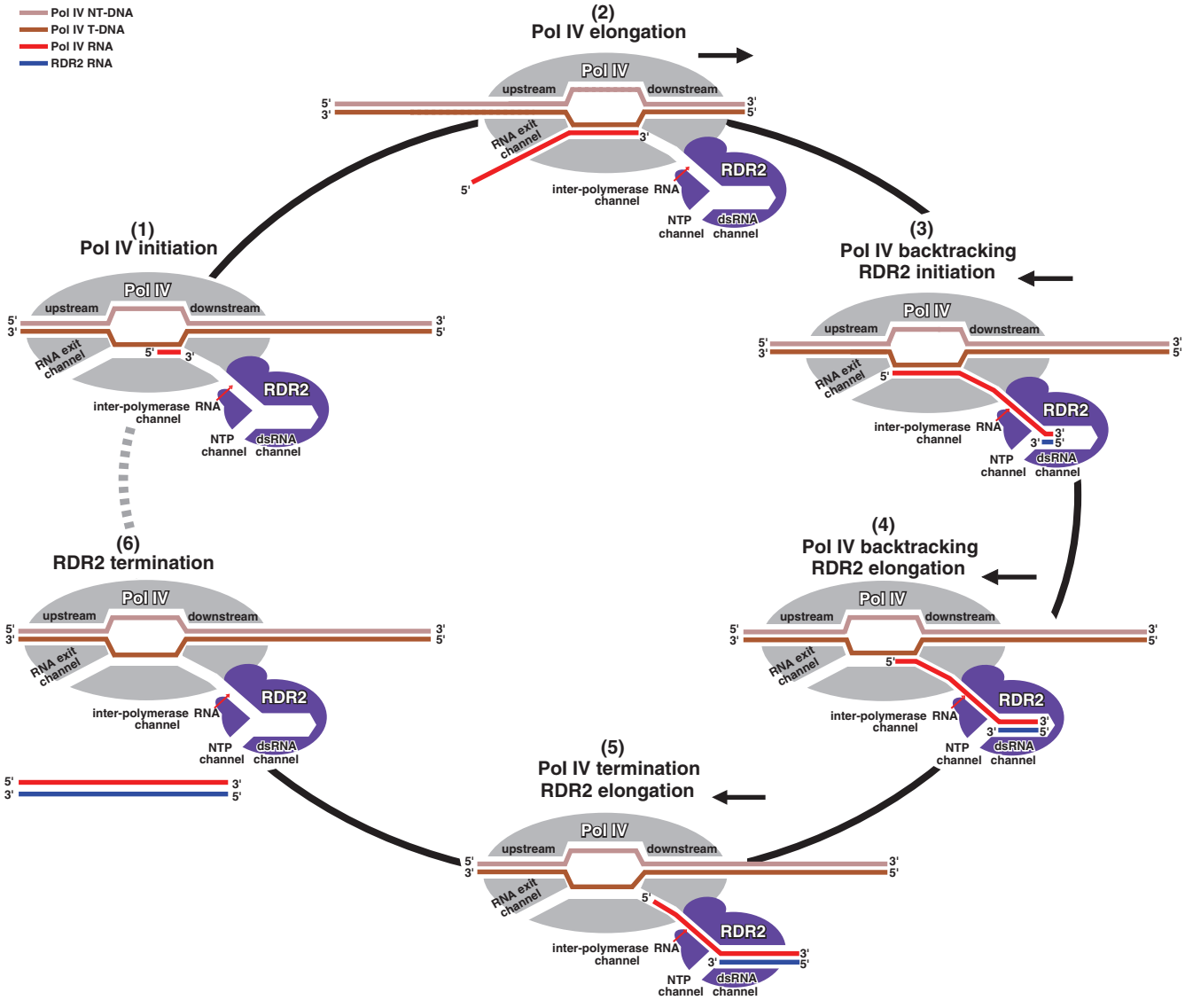


Fig. 6. Backtracking-triggered RNA channeling model for dsRNA synthesis by Pol IV-RDR2. Black arrows indicate the direction of Pol IV–RDR2 movement on DNA.

methylation. By contrast, P_{NRPD1}::*NRPD1* (ΔFHT; M5) *nrpd1* exhibited the same low-DNA methylation level as the *nrpd1* mutant. Moreover, the 24-nt siRNAs associated with the RdDM loci were substantially reduced in P_{NRPD1}::*NRPD1* (ΔFHT; M5) *nrpd1* plants, compared with wild type and P_{NRPD1}::*NRPD1 nrpd1* (fig. S12). These results demonstrate that the Pol IV–RDR2 interaction is essential for production of the dsRNA precursors of 24-nt siRNA and the subsequent DNA methylation.

On the basis of our cryo-EM structures and previous results from biochemical and genetic studies, we propose a backtracking-triggered RNA channeling model of dsRNA synthesis by the Pol IV–RDR2 complex (Fig. 6): (i) Pol IV–RDR2 is recruited to genomic RdDM loci by SHH1- and CLASSY-family proteins to initiate synthesis of Pol IV RNA (*21*–*25*); (ii) Pol IV moves forward and elongates Pol IV RNA to a length of 30 to 40 nt (*6*, *7*); (iii) Pol IV backtracks, and Pol IV RNA is threaded through the interpolymerase RNA channel into the RDR2 active site, where Pol IV RNA is trapped and used as the template for RDR2 RNA synthesis; (iv) RDR2 elongates RDR2 RNA resulting in further Pol IV backtracking; (v) RDR2 pulls Pol IV RNA out of the Pol IV active center and releases Pol IV RNA, resulting in Pol IV termination; and (vi) RDR2 releases dsRNA from its active site. After completing a transcription cycle, Pol IV returns to the position where the Pol IV RNA synthesis initiates and presumably could initiate another round of Pol IV RNA synthesis to amplify small RNA signals for local or long-distance action (*47*).

We predict that the lack of TFIIS interaction renders Pol IV inefficient in cleavage of backtracked RNA during elongation and ensures that completely backtracked RNA is delivered into the RDR2 active center. Pol IV backtracking can be triggered by increased dwell time in the pretranslocation state caused by dsDNA unwinding incapability (*20*), torsion stress posed by positively supercoiled downstream DNA (*48*), and/or nucleosome roadblocks (*49*, *50*). A recent report revealed that RDR2 RNA synthesis cannot proceed unless Pol IV unwinds a downstream dsDNA of >17 bp in length (*20*). This finding supports the minimal requirement of backtracked nascent Pol IV RNA to reach the RDR2 active center in our model, suggesting a possible role of dsDNA unwinding in Pol IV RNA backtracking. Further structural and biochemical studies are required to elucidate how Pol IV backtracking is achieved in vivo.

The structures of Pol IV–RDR2 complexes provide atomic details for such an unprecedented two-RNA-polymerase assembly (*18*). The interpolymerase RNA channel revealed

in our work provides structural evidence for the RNA channeling hypothesis (*20*). However, contradictory to previous assumptions that Pol IV RNA has to be completely released from the Pol IV RNA exit channel in a 5′-to-3′ direction to expose the 3′ terminus for RDR2 engagement (*20*, *51*), we found that the Pol IV RNA is threaded inside of the two-RNA-polymerase complex through the inter-polymerase RNA channel in a 3′-to-5′ direction to the RDR2 active center, where the 3′ terminus of Pol IV RNA is directly captured and used as the template for RDR2 RNA synthesis.

The backtracking-triggered RNA channeling model suggests that RDR2 pulls and releases Pol IV RNA in a 3′ to 5′ direction from the Pol IV secondary channel, and thus represents a previously undescribed means of transcription termination by Pol IV, a dsRNA-induced termination (Fig. 6). This termination mode is distinct from all previous transcription termination models (*52*, *53*). The dsRNA-induced termination model predicts that Pol IV RNA production should be hindered if Pol IV is uncoupled from RDR2; this prediction is supported by previous in vivo RNA-seq results which showed that Pol IV RNA production is impaired in RDR2 mutants (*6*, *8*, *20*), and by a previous in vitro transcription study which showed that Pol IV is unable to terminate properly if RDR2 RNA transcripts are not detected (*20*).

REFERENCES AND NOTES

1. R. G. Roeder, W. J. Rutter, *Nature* **224**, 234–237 (1969).
2. Y. Onodera *et al.*, *Cell* **120**, 613–622 (2005).
3. A. J. Herr, M. B. Jensen, T. Dalmay, D. C. Baulcombe, *Science* **308**, 118–120 (2005).
4. T. Kanno *et al.*, *Nat. Genet.* **37**, 761–765 (2005).
5. D. Pontier *et al.*, *Genes Dev.* **19**, 2030–2040 (2005).
6. J. Zhai *et al.*, *Cell* **163**, 445–455 (2015).
7. T. Blevins *et al.*, *eLife* **4**, e09591 (2015).
8. S. Li *et al.*, *Genome Res.* **25**, 235–245 (2015).
9. A. T. Wierzbicki, J. R. Haag, C. S. Pikaard, *Cell* **135**, 635–648 (2008).
10. A. T. Wierzbicki, T. S. Ream, J. R. Haag, C. S. Pikaard, *Nat. Genet.* **41**, 630–634 (2009).
11. M. El-Shami *et al.*, *Genes Dev.* **21**, 2539–2544 (2007).
12. M. A. Matzke, T. Kanno, A. J. Matzke, *Annu. Rev. Plant Biol.* **66**, 243–267 (2015).
13. H. Zhang, Z. Lang, J. K. Zhu, *Nat. Rev. Mol. Cell Biol.* **19**, 489–506 (2018).
14. Ö. Deniz, J. M. Frost, M. R. Branco, *Nat. Rev. Genet.* **20**, 417–431 (2019).
15. R. A. Mosher *et al.*, *Nature* **460**, 283–286 (2009).
16. T. S. Ream *et al.*, *Mol. Cell* **33**, 192–203 (2009).
17. J. Luo, B. D. Hall, *J. Mol. Evol.* **64**, 101–112 (2007).
18. J. R. Haag *et al.*, *Mol. Cell* **48**, 811–818 (2012).
19. M. Marasco, W. Li, M. Lynch, C. S. Pikaard, *Nucleic Acids Res.* **45**, 11315–11326 (2017).
20. J. Singh, V. Mishra, F. Wang, H. Y. Huang, C. S. Pikaard, *Mol. Cell* **75**, 576–589.e5 (2019).
21. J. A. Law *et al.*, *Nature* **498**, 385–389 (2013).
22. M. Zhou, A. M. S. Palanca, J. A. Law, *Nat. Genet.* **50**, 865–873 (2018).
23. H. Zhang *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 8290–8295 (2013).
24. L. M. Smith *et al.*, *Plant Cell* **19**, 1507–1521 (2007).
25. J. A. Law, A. A. Vashisht, J. A. Wohlschlegel, S. E. Jacobsen, *PLOS Genet.* **7**, e1002195 (2011).
26. R. Lister *et al.*, *Cell* **133**, 523–536 (2008).
27. J. Li *et al.*, *Nat. Plants* **6**, 661–674 (2020).
28. Y. Ogawa *et al.*, *Plant Cell Physiol.* **49**, 242–250 (2008).
29. R. H. Ebright, *J. Mol. Biol.* **304**, 687–698 (2000).
30. P. S. Salgado *et al.*, *PLOS Biol.* **4**, e434 (2006).
31. P. Cramer, D. A. Bushnell, R. D. Kornberg, *Science* **292**, 1863–1876 (2001).
32. D. Wang, D. A. Bushnell, K. D. Westover, C. D. Kaplan, R. D. Kornberg, *Cell* **127**, 941–954 (2006).
33. X. Chen *et al.*, *Science* **372**, eaba8490 (2021).
34. S. Aibara, S. Schilbach, P. Cramer, *Nature* **594**, 124–128 (2021).
35. M. Wind, D. Reines, *BioEssays* **22**, 327–336 (2000).
36. A. C. Cheung, P. Cramer, *Nature* **471**, 249–253 (2011).
37. D. Wang *et al.*, *Science* **324**, 1203–1206 (2009).
38. H. Kettenberger, K. J. Armache, P. Cramer, *Mol. Cell* **16**, 955–965 (2004).
39. T. Dalmay, A. Hamilton, S. Rudd, S. Angell, D. C. Baulcombe, *Cell* **101**, 543–553 (2000).
40. E. Nudler, *Cell* **149**, 1438–1445 (2012).
41. A. E. Sluder, D. H. Price, A. L. Greenleaf, *J. Biol. Chem.* **263**, 9917–9925 (1988).
42. A. Fukudome *et al.*, *bioRxiv* [Preprint] 2021.2008.2031.458351 (2021).
43. A. L. Gnatt, P. Cramer, J. Fu, D. A. Bushnell, R. D. Kornberg, *Science* **292**, 1876–1882 (2001).
44. C. O. Barnes *et al.*, *Mol. Cell* **59**, 258–269 (2015).
45. L. Li, V. Molodtsov, W. Lin, R. H. Ebright, Y. Zhang, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 5801–5809 (2020).
46. S. Sainsbury, J. Niesser, P. Cramer, *Nature* **493**, 437–440 (2013).
47. A. Molnar *et al.*, *Science* **328**, 872–875 (2010).
48. J. Ma, L. Bai, M. D. Wang, *Science* **340**, 1580–1583 (2013).
49. L. Farnung, S. M. Vos, P. Cramer, *Nat. Commun.* **9**, 5432 (2018).
50. H. Ehara *et al.*, *Science* **363**, 744–747 (2019).
51. V. Mishra *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2019276118 (2021).
52. N. J. Proudfoot, *Science* **352**, aad9926 (2016).
53. O. Porrua, D. Libri, *Nat. Rev. Mol. Cell Biol.* **16**, 190–202 (2015).

ACKNOWLEDGMENTS

We thank T. Xu for the *A. thaliana* T87 cell line, L. Kong, F. Wang, G. Li, and J. Duan at the cryo-EM center of the National Center for Protein Science Shanghai; B. Zhu and X. Li at the center for Biological Imaging of the CAS Institute of Biophysics; S. Chang at the cryo-EM center of Zhejiang University; and X. Ma at the cryo-EM center of SUST for assistance with data collection. **Funding:** J.W. received National Natural Science Foundation of China Grant 31788103 and Strategic Priority Research Program of CAS grant XDB27030101; Y.Z. received the Shanghai Pilot Program for Basic Research – Chinese Academy of Science, Shanghai Branch, National Natural Science Foundation of China grant 31822001, and Leading Science Key Research Program of CAS grant QYZDB-SSW-SMC005. **Author contributions:** Conceptualization: Y.Z. and J.W. Pol IV preparation: K.H., C.Z., J.G., H.Z., Z.G., and W.M. Structure determination: K.H., Y.F., C.F., X.W., and L.Y. Transcription assay: C.F. and K.H. RNA and DNA sequencing: Z.X. and K.H. Funding acquisition: Y.Z., J.W., and Y.F. Project administration: Y.Z., J.W., and Y.F. Supervision: Y.Z., J.W., and Y.F. Writing – original draft: Y.Z. and J.W. Writing – review and editing: Y.Z., J.W., Y.F., K.H., X.W. and C.F. **Competing interests:** Authors declare that they have no competing interests. **Data and materials availability:** The cryo-EM map and coordinates were deposited in Protein Data Bank and Electron Microscopy Data Bank (Pol IV–RDR2 holoenzyme: 7EU1 and EMD-31306; Pol IV–RDR2 bTEC: 7EU0 and EMD-31305). The small RNA-seq and bisulfite-seq data (BioProject PRJCA005428) were deposited in the Beijing Institute of Genomics Data Center (https://ngdc.cncb.ac.cn/).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abj9184
Materials and Methods
Figs. S1 to S12
Tables S1 to S3
References (*54*–*66*)
MDAR Reproducibility Checklist
Movies S1 to S5

14 June 2021; accepted 9 November 2021
10.1126/science.abj9184

CORONAVIRUS

An oral SARS-CoV-2 M^{pro} inhibitor clinical candidate for the treatment of COVID-19

Dafydd R. Owen[1]*, Charlotte M. N. Allerton[1], Annaliesa S. Anderson[2], Lisa Aschenbrenner[3], Melissa Avery[3], Simon Berritt[3], Britton Boras[4], Rhonda D. Cardin[2], Anthony Carlo[3], Karen J. Coffman[3], Alyssa Dantonio[3], Li Di[3], Heather Eng[3], RoseAnn Ferre[4], Ketan S. Gajiwala[4], Scott A. Gibson[5], Samantha E. Greasley[4], Brett L. Hurst[5], Eugene P. Kadar[3], Amit S. Kalgutkar[1], Jack C. Lee[3], Jisun Lee[3], Wei Liu[4], Stephen W. Mason[2]†, Stephen Noell[3], Jonathan J. Novak[3]‡, R. Scott Obach[3], Kevin Ogilvie[3], Nandini C. Patel[1], Martin Pettersson[1]§, Devendra K. Rai[2], Matthew R. Reese[3], Matthew F. Sammons[1], Jean G. Sathish[2], Ravi Shankar P. Singh[1], Claire M. Steppan[3], Al E. Stewart[4], Jamison B. Tuttle[1], Lawrence Updyke[1], Patrick R. Verhoest[1], Liuqing Wei[3], Qingyi Yang[1], Yuao Zhu[2]

The worldwide outbreak of COVID-19 caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has become a global pandemic. Alongside vaccines, antiviral therapeutics are an important part of the healthcare response to countering the ongoing threat presented by COVID-19. Here, we report the discovery and characterization of PF-07321332, an orally bioavailable SARS-CoV-2 main protease inhibitor with in vitro pan-human coronavirus antiviral activity and excellent off-target selectivity and in vivo safety profiles. PF-07321332 has demonstrated oral activity in a mouse-adapted SARS-CoV-2 model and has achieved oral plasma concentrations exceeding the in vitro antiviral cell potency in a phase 1 clinical trial in healthy human participants.

Human coronavirus infections are common, with at least four examples (229E, NL63, OC43, and HKU1) now considered endemic (*1*). However, the emergence within the past 20 years of severe acute respiratory syndrome coronavirus 1 (SARS-CoV-1), Middle East respiratory syndrome coronavirus (MERS-CoV), and SARS-CoV-2 as novel human coronaviruses has signaled the substantial threat potential of this viral class. The catastrophic SARS-CoV-2 outbreak of 2019 has resulted

in 235 million confirmed cases of COVID-19 causing >5.2 million deaths globally as of December 2021. SARS-CoV-2 is a highly infectious, RNA beta-coronavirus that can cause life-threatening viral pneumonia in the most serious cases. Although effective COVID-19 vaccines have been developed within unprecedented timelines, a large number of people are either unable due to preexisting medical conditions or unwilling to be vaccinated, and global access challenges remain. Limited therapeutic options are available to those who are infected. Oral SARS-CoV-2–specific therapeutics are urgently needed to prevent more severe disease, hospitalization, and death. Treatment may also reduce the period of infectivity. Repurposing of approved drugs in the search for small-molecule antiviral agents that target SARS-CoV-2 has thus far been minimally effective (*2*, *3*). Viral RNA–dependent RNA polymerase inhibitors such as molnupiravir and AT-527 are currently undergoing clinical trials for the treatment of COVID-19 (*4*, *5*).

The SARS-CoV-2 genome encodes two polyproteins, pp1a and pp1ab, and four structural proteins (*6*, *7*). The polyproteins are cleaved by the critical SARS-CoV-2 main protease (M^{pro}, also referred to as 3CL protease) at 11 different sites to yield shorter, nonstructural proteins vital to viral replication (*8*, *9*). The coronavirus M^{pro} is a three-domain cysteine protease that features a Cys145-His41 catalytic dyad located in the cleft between domains I and II. Several common features are shared among M^{pro} substrates, including the presence of a Gln residue at P1 [using Schechter-Berger nomenclature (*10*)]. No known human cysteine protease cleaves after Gln, thus offering potential selectivity for this viral target over the human proteome (*11*–*13*). Viral proteases are tractable targets for small-molecule oral therapies in the treatment of HIV and HCV (*14*, *15*). Moreover, a recent report has demonstrated oral activity of an M^{pro} inhibitor in a transgenic mouse model of SARS-CoV-2 infection (*16*). Given the pivotal role of SARS-CoV-2 M^{pro} in viral replication, its potential for mechanistic safety, and the expected lack of spike protein variant resistance challenges, SARS-CoV-2 M^{pro} inhibition represents an attractive small-molecule approach for an oral antiviral therapy to treat COVID-19.

[1]Pfizer Worldwide Research, Development & Medical, Cambridge, MA 02139, USA. [2]Pfizer Worldwide Research, Development & Medical, Pearl River, NY 10965, USA. [3]Pfizer Worldwide Research, Development & Medical; Groton, CT 06340, USA. [4]Pfizer Worldwide Research, Development & Medical, La Jolla, CA 92121, USA. [5]Institute for Antiviral Research, Department of Animal, Dairy, and Veterinary Sciences, Utah State University; Logan, UT 84322, USA.
***Corresponding author. Email: Dafydd.owen@pfizer.com**
†Present address: Janssen Biopharma; South San Francisco, CA 94080, USA.
‡Present address: Praxis Precision Medicines; Cambridge, MA 02142, USA.
§Present address: GRT Therapeutics; Cambridge, MA 02142, USA.

An effort to identify inhibitors of the SARS-CoV-1 M^{pro} in response to the 2002 SARS outbreak led to the identification of PF-00835231 (**1**; Fig. 1) as a potent inhibitor of recombinant SARS-CoV-1 M^{pro} in a fluorescence resonance energy transfer (FRET)–based substrate cleavage assay (*17*). PF-00835231 also demonstrated potent inhibition [inhibition constant (K_i) = 0.271 nM] of recombinant SARS-CoV-2 M^{pro}, which is expected given that the SARS-CoV-1 and SARS-CoV-2 M^{pro} share 100% sequence homology across their respective substrate-binding sites (*18*). Antiviral activity against SARS-CoV-2 was also observed with PF-00835231 [half-maximal effective concentration (EC_{50}) of 231 nM] by monitoring the cytopathic effect (CPE) in epithelial Vero E6 cells. Because Vero E6 cells express high levels of the efflux transporter P-glycoprotein, for which **1** is a substrate (*17*, *19*), antiviral assays in this cell line were conducted in the presence of the P-glycoprotein efflux inhibitor CP-100356 (*20*). TMPRSS2 expression was not detected in the Vero E6 cells by quantitative reverse transcription polymerase chain reaction. PF-07304814, the phosphate prodrug form of PF-00835231, is currently under investigation as an intravenous treatment option for COVID-19 in hospitalized patients (*20*).

To improve upon the low passive absorptive permeability (P_{app} < 0.207 × 10^{-6} cm/s) (*21*) and poor oral absorption of **1** in animals, we aimed to remove the hydrogen bond donor (HBD) of the P1′ α-hydroxymethyl ketone moiety in **1** (Fig. 2A) because increased HBD count has been shown to be correlated with poor oral bioavailability (*22*). To this end, we pursued two functional groups precedented as covalent warheads for cysteine proteases in parallel: nitriles (*23*, *24*) and benzothiazol-2-yl ketones (*25*, *26*). The nitrile compound **2** demonstrated an increase in rat oral absorption [oral bioavailability (F) = 7.6% and fraction of oral dose absorbed from the gastrointestinal tract ($F_a \times F_g$) = 38%] (*27*) while maintaining reasonable metabolic stability [intrinsic clearance (CL_{int})] against oxidative metabolism in human liver microsomes (HLMs) (*28*) relative to **1** (Fig. 1). However, the in vitro FRET M^{pro} potency (K_i = 27.7 nM) and SARS-CoV-2 antiviral activity (EC_{50} = 1364 nM) of **2** was inferior to **1**. Introduction of a 6,6-dimethyl-3-azabicyclo[3.1.0]hexane as a cyclic leucine mimetic at P2 (Fig. 2B) removed the HBD from the P2/P3 amide linkage. Analog **3**, resulting from the combination of this cyclic P2 fragment with a P1′ benzothiazolyl ketone, displayed high P_{app} (10.3 × 10^{-6} cm/s). The reduced biochemical SARS-CoV-2 M^{pro} inhibitory potency of **3** (K_i = 230 nM) relative to other reported benzothiazole-2-yl SARS-CoV-2 M^{pro} inhibitors (*29*) containing leucine P2 groups can be rationalized from the binding mode observed for **3** (Fig. 2C). Although the 6,6-dimethyl-3-azabicyclo[3.1.0]hexane effectively fills the lipophilic S2 pocket formed by Met49, Met169, His41, and Gln189, productive hydrogen bonding of the ligand backbone to Gln189 is no longer possible (Fig. 2C). The inferior SARS-CoV-2 M^{pro} potency and the high CL_{int} (337 μl/min/mg) precluded further investments in compound **3**. Similar to **1**, the P3 indole of **3** does not protrude into the S3 pocket (Fig. 2, A and C). To better occupy the S3 pocket, we introduced branched, acyclic P3 groups. The methanesulfonamide in compound **4** extends underneath Gln189, productively engaging P3 pocket residues and achieving improved hydrogen-bonding interactions with the Glu166 backbone (Fig. 2D) relative to **1** and **3**. Compound **4** demonstrated improved SARS-CoV-2 M^{pro} biochemical potency (K_i = 7.93 nM), Vero E6 antiviral activity (EC_{50} = 909 nM), and HLM CL_{int} (127 μl/min/mg) relative to **3** (Fig. 1). Examination of the rat pharmacokinetics of **4** also revealed improvements in oral absorption (F = 10%, $F_a \times F_g$ = 84%) (Fig. 1). An effort to identify alternate P3 capping groups to sulfonamide led to trifluoroacetamide **5**. Compound **5** exhibited comparable biochemical potency (K_i = 12.1 nM) to **4** but with greatly improved SARS-CoV-2 Vero E6 antiviral activity (EC_{50} = 85.3 nM and P_{app} = 13.1 × 10^{-6} cm/sec) as well as increased metabolic stability in HLMs (CL_{int} = 30.3 μl/min/mg) (Fig. 1). **5** also shows greatly improved oral pharmacokinetics in both rats (F = 33%, $F_a \times F_g$ = 100%) (Fig. 1) and monkeys (F = 7.9%, $F_a \times F_g$ = 66%) (table S1). Introduction of the P1′ nitrile to this scaffold led to the identification of the clinical candidate PF-07321332 (**6**; Fig. 1). Compound **6** is a potent inhibitor of SARS-CoV-2 M^{pro} biochemical activity (K_i = 3.11 nM) with improved Vero E6 antiviral activity (EC_{50} = 74.5 nM) relative to compounds **2** to **4**. Compound **6** displays a similar binding mode (Fig. 2E) to compound **4**. The P1′ nitrile of **6** forms a reversible covalent thioimidate adduct with the catalytic Cys145 (Fig. 2F). Reversibility of M^{pro} inhibition by **6** was demonstrated upon incubation of SARS-CoV-2 M^{pro} (2 mM) with 2 mM of either **6** or an irreversible M^{pro} inhibitor (compound **7**; Fig. 3A) (*17*) for 30 min and monitoring M^{pro} activity after 100-fold dilution of the incubation mixtures. No recovery of activity was observed after M^{pro} incubation with **7**. The recovery of >50% M^{pro} activity after incubation with **6** indicates that inhibition of SARS-CoV-2 M^{pro} is reversible (Fig. 3A). We selected the nitrile compound **6** (named PF-07321332) over compound **5** as the clinical candidate based on ease of synthetic scale-up, enhanced solubility that allowed for a simple formulation vehicle in support of preclinical toxicology, and reduced propensity for epimerization at the P1 stereocenter.

PF-07321332 demonstrated potent inhibition in FRET M^{pro} assays representing M^{pro}

from all coronavirus types known to infect humans (*6*, *7*, *30*), including beta-coronaviruses (SARS-CoV-2, SARS-CoV-1, HKU1, OC43, and MERS-CoV) as well as alpha-coronaviruses (229E and NL63) (Fig. 3B and table S2). No inhibitory effects were noted against several mammalian cysteine (caspase 2, cathepsin B, and cathepsin L), serine (chymotrypsin, elastase, and thrombin) and aspartyl (cathepsin D) proteases at the highest concentration tested (100 μM) of PF-07321332 (table S3). This was also the case for HIV-1 protease, a viral aspartyl protease (table S3).

The in vitro antiviral activity of PF-07321332 was also evaluated in two physiologically relevant

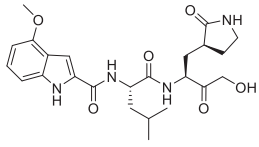
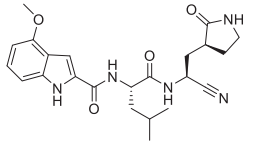
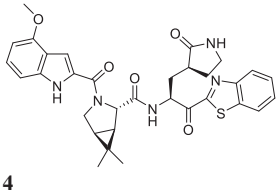
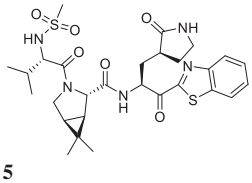
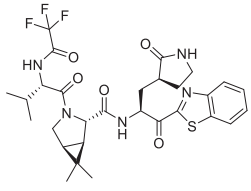
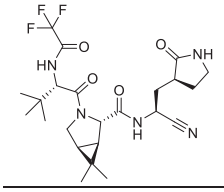
Number/Structure	SARS-CoV2 M^{pro} K_i (nM)*	VeroE6-enACE2 CPE EC_{50} (nM)†	MDCK-LE P_{app} (x 10^{-6} cm/sec)‡	HLM CL_{int} (μl/min/mg)§	Rat CL_p (mL/min/kg)¶	Oral F (%)¶#	F_a x F_g (%)¶††
1 (PF-00835231) 	0.271 (0.155 – 0.471, n=6)	231 (158 – 338, n=8)	< 0.207 ± 0.048 (n=6)	7.47 ± 0.88	27.0 ± 3.1	1.4 ± 0.8	3.3
2 	27.7 (18.4 – 41.7, n=5)	1364 (860 – 2164, n=15)	0.945 ± 0.281 (n=6)	34.4 ± 0.7	39.3 (37.0, 41.5)	7.6 (7.4, 7.8)	38
3 	230 (181 – 292, n=4)	5593 (3457 – 9051, n=8)	10.3 ± 2.4 (n=6)	337 ± 9	N.D.	N.D.	N.D.
4 	7.93 (3.62 – 17.4, n=5)	909 (557 – 1482, n=14)	1.56 ± 0.38 (n=6)	127 ± 3	42.9 (38.2, 47.6)	10 (7.5, 13)	84
5 	12.1 (8.05 – 18.1, n=7)	85.3 (76.5– 95.2, n=36)	13.1 ± 2.0 (n=8)	30.3 ± 0.6	31.0 (30.6, 31.4)	33 (33, 34)	100
6 (PF-07321332) 	3.11 (1.47 – 6.59, n=6)	74.5 (66.5 – 83.4, n=20)	1.71 ± 0.28 (n=4)	24.5 ± 0.2	27.2 (22.5, 31.9)	50 (30, 71), 34 ±19**	95, 65**

Fig. 1. In vitro and in vivo parameters optimized in identifying oral SARS-CoV-2 M^{pro} inhibitors. *K_i values were fit to the Morrison equation with substrate, K_m, and M^{pro} concentration parameters fixed to values described in the supplementary materials. Data are geometric mean values, with 95% confidence interval (CI) values and replicate numbers in parentheses. †EC_{50} values were calculated using data normalized to controls within the assay and fit to a four-parameter logistic curve fit (see the supplementary materials for details). Data are shown as geometric means, with 95% CI values and replicate numbers in parentheses. ‡P_{app} from apical to basolateral direction was determined in Madin-Darby canine kidney-low efflux (MDCK-LE) cells (*21*). §CL_{int} refers to total intrinsic clearance obtained from scaling of half-lives of test compounds in NADPH-supplemented HLMs (*28*). Incubations were conducted on a single day in triplicate. ¶Pharmacokinetic parameters were calculated from plasma concentration–time data and are reported as mean values (n = 2 to 3 male Wistar-Han rats/dosing route) (see the supplementary materials for details). #Oral pharmacokinetics studies were conducted in the fed state. F is defined as the dose-normalized AUC after oral administration divided by the dose-normalized AUC after intravenous administration. **Crystalline **6** was orally administered in anhydrous (form 1) as well as anhydrous methyl-tert-butyl ether co-solvate form. ††$F_a \times F_g$ was estimated using the equation $F_a \times F_g = F/(1 - CL_{blood}/Q)$ (*27*) (see the supplementary materials for details). N.D., not determined.

cellular systems: human adenocarcinoma–derived alveolar basal epithelial (A549) cells constitutively expressing ACE2 (*19*) and differentiated normal human bronchial epithelial (dNHBE) cells (*31*). In contrast to Vero E6 cells, A549 and dNHBE cells do not overly express P-glycoprotein, and therefore co-dosing with a P-glycoprotein inhibitor was not necessary to gauge antiviral activity in these cell lines (*19*). PF-07321332 inhibited SARS-CoV-2 replication as assessed using a nanoluciferase reporter virus in A549-ACE2 cells with EC_{50} and EC_{90} values of 77.9 and 215 nM, respectively, with no cytotoxicity detected at concentrations up to 3 μM (Fig. 3C and table S4). Treatment of dNHBE cells with varying concentrations of PF-07321332 for 3 days led to inhibition of SARS-CoV-2 viral replication, with EC_{50} and EC_{90} values of 61.8 and 181 nM, respectively (Fig. 3C), as monitored by titration of virus harvested from the apical compartment using a 50% cell culture infective dose ($CCID_{50}$) assay in Vero76 cells. Increasing the duration of the dNHBE study to 5 days resulted in viral replication being inhibited, with EC_{50} and EC_{90} values of 32.6 and 56.1 nM, respectively (Fig. 3C). Because optimal therapeutic efficacy with marketed viral protease inhibitors and other antiviral agents is generally achieved when the minimum systemic unbound plasma concentration (C_{min}) of inhibitor is maintained above cellular antiviral EC_{90} (*32*), we selected the more conservative day 3 EC_{90} value of 181 nM PF-07321332 in the dNHBE assay as the C_{min} to be maintained when predicting efficacy in animal models and in the clinical setting. We further evaluated the in vitro cellular antiviral activity of PF-07321332 against SARS-CoV-1, MERS-CoV, and human coronavirus 229E using CPE assays. PF-07321332 demonstrated potent antiviral activity against SARS-CoV-1 (EC_{90} = 317 nM), MERS-CoV (EC_{90} = 351 nM), and 229E (EC_{90} = 620 nM) in their respective cellular assays (Fig. 3D and table S5).

We evaluated the in vivo antiviral activity of PF-07321332 in a mouse-adapted SARS-CoV-2 (SARS-CoV-2 MA10) model (*33*). Intranasal infection of BALB/c mice with SARS-CoV-2 MA10 led to ~10% body weight loss and minimal mortality in 10-week-old mice. As shown in Fig. 4A, after infection with SARS-CoV-2 MA10, mice treated twice daily with PF-07321332 (at both 300 and 1000 mg/kg doses) were protected from weight loss compared with vehicle-treated mice. At 4 days after infection, mice were sacrificed and lung viral titers were evaluated in $CCID_{50}$ assays. Infected animals in the placebo group (n = 12, two independent studies) had robust infection in the lungs (mean lung titer of $\log_{10}$ 4.93 ± 0.140 $CCID_{50}$/ml SARS-CoV-2 MA10) (Fig. 4B), whereas virus levels in mice treated with PF-07321332 were significantly reduced (mean lung titers of $\log_{10}$ 3.53 ± 0.187 and $\log_{10}$ 3.02 ± 0.423 $CCID_{50}$/ml for the 300 and 1000 mg/kg PF-07321332-treated groups, respectively). In a satellite group of uninfected mice, the 300 mg/kg twice daily (BID) dose of PF-07321332 used in the mouse-adapted viral in vivo efficacy study maintained

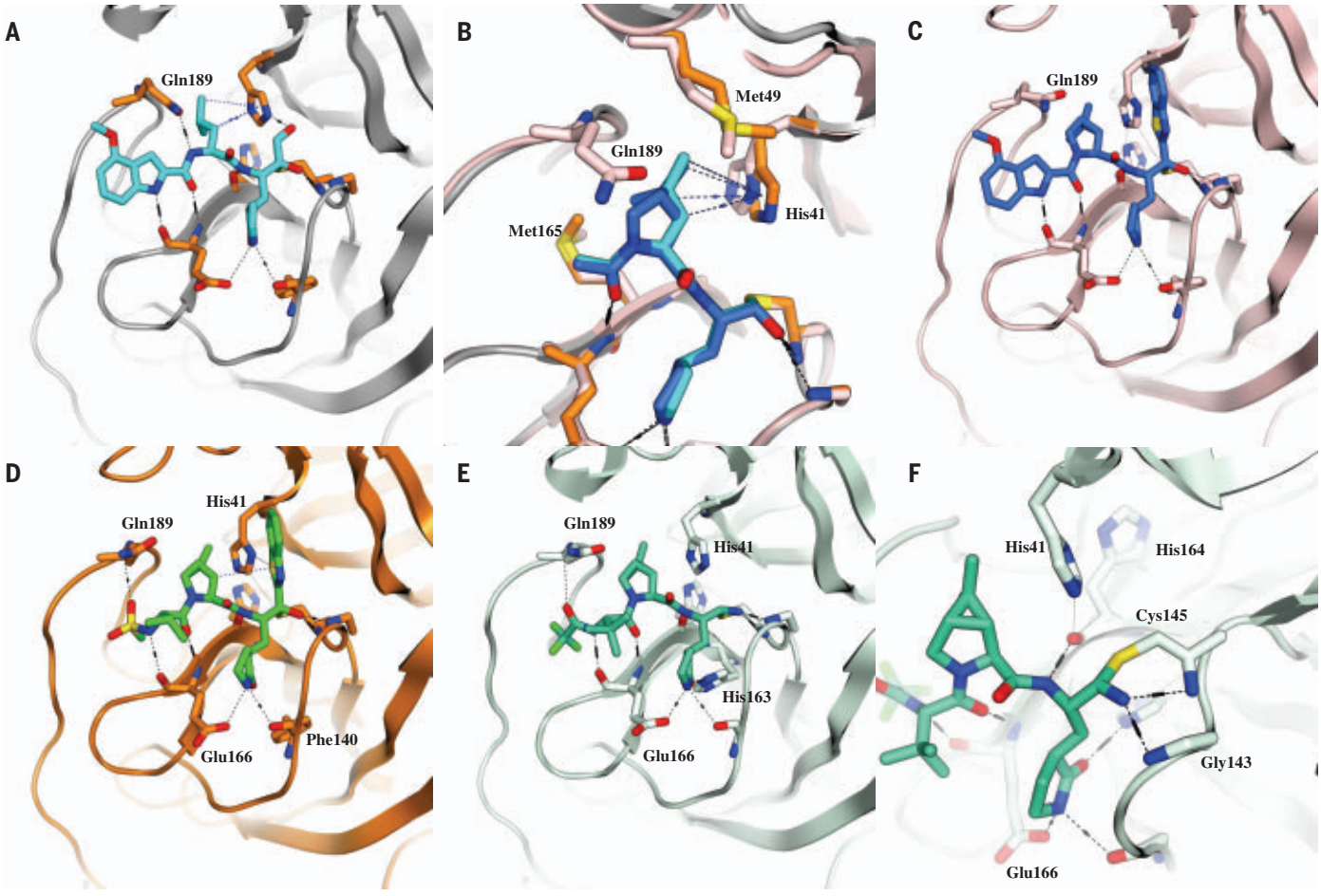
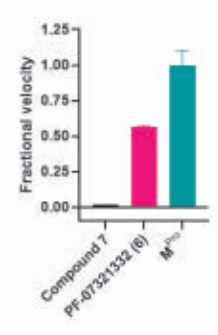
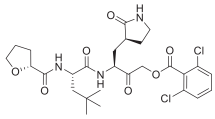


Fig. 2. SARS-CoV-2 M^{pro} structural biology. (**A**) Co-crystal structure of PF-00835231 (**1**) with SARS-CoV-2 M^{pro}. Key interactions are indicated. (**B**) Modeled overlap of dimethyl-bicyclo[3.1.0] proline from compound **3** (blue) as a mimic of P2 leucine residue (cyan) found in the viral polyprotein substrate and **1**. This tolerated P2 change eliminates an H-bond donor from resulting inhibitors. (**C**) Compound **3** effectively fills the lipophilic S2 pocket formed by Met49, Met165, and His41, but productive hydrogen bonding to Gln189 is no longer possible. (**D**) Compound **4** with optimized acyclic P3 group and restored Gln189 interaction. (**E**) SARS-CoV-2 M^{pro}–bound crystal structure of clinical candidate PF-07321332 (**6**). (**F**) A reversible covalent Cys145 adduct is formed with the nitrile substituent in compound **6**.

A

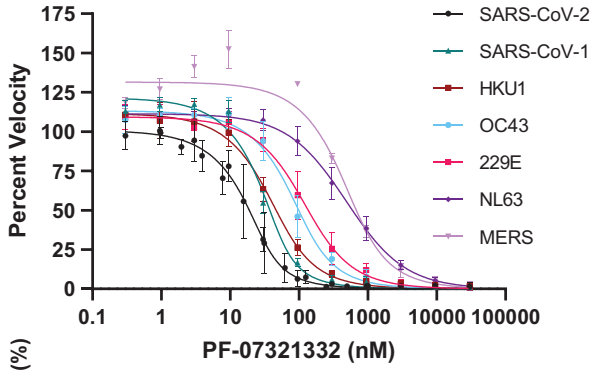




7

SARS-CoV2 M^{pro} FRET K_i = 2.5 nM

B



C

Cell line	Endpoint (Assay)	EC_{50} nM (95% CI)	EC_{90} nM (95% CI)	n
Vero E6 + 2mM CP-100356	Cell Viability (CPE)	74.5 (66.5 to 83.4)	155 (138 to 173)	20
A549-ACE2	Viral Replication (Nanoluciferase Reporter Virus)	77.9 (50.5 to 120)	215 (146 to 317)	28
dNHBE day 3	Viral Titer ($CCID_{50}$)	61.8 (32.4 to 118)	181 (76.9 to 425)	3
dNHBE day 5	Viral Titer ($CCID_{50}$)	32.6 (10.2 to 104)	56.1 (19.2 to 164)	3

D

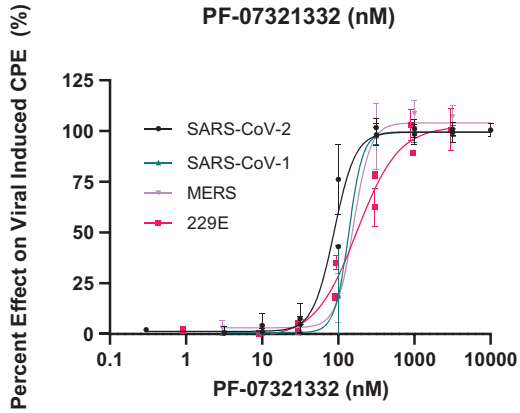


Fig. 3. PF-07321332 biochemical and antiviral activity. (**A**) PF-07321332 is a reversible inhibitor of SARS-CoV-2 M^{pro} as demonstrated by recovery of enzymatic activity after a 100-fold dilution of the enzyme inhibitor complex. Compound **7** (PF-00956378), an irreversible inhibitor, was included as a control. Data are representative of three independent experiments. (**B**) PF-07321332 is a potent inhibitor of the proteolytic activity of SARS-CoV-2 M^{pro} as well as related coronaviruses in FRET assays. Data shown are the mean ± SD from three independent experiments. (**C**) PF-07321332 demonstrates potent SARS-CoV-2 antiviral cellular activity. PF-07321332 inhibited SARS-CoV-2–induced CPE in Vero E6 cells enriched for ACE2. A P-glycoprotein inhibitor, CP-100356 (efflux inhibitor, EI), was added at 2 μM to inhibit the P-glycoprotein–mediated efflux of PF-07321332. In Vero E6 cells with no EI, the EC_{50} of PF-07321332 [95% CI was 4.48 μM (3.55 to 5.65 μM); N = 8, n = 20]. Cytotoxicity of PF-07321332 was evaluated in noninfected cells, and the $CCID_{50}$ was >100 μM (table S4). PF-07321332 inhibits SARS-CoV-2 replication in A549 cells expressing ACE2 in four independent experiments. The $CCID_{50}$ in noninfected A549 cells was >3 μM (table S4). In dNHBE cells, PF-07321332 decreased SARS-CoV-2 viral replication (N = 3). Data are shown as the geometric mean and 95% CI. (**D**) PF-07321332 demonstrates pan-coronavirus antiviral activity. PF-07321332 inhibition in viral-induced CPE assays: SARS-CoV-1 in Vero E6 cells (in the presence of 2 μM EI CP-100356), h-CoV-229E in MRC-5 cells, and MERS-CoV in Vero 81 cells (in the presence of 1 μM EI CP-100356). Data are shown as mean ± SD. $CCID_{50}$ values were determined in all assays to be >100 μM (table S4).

C_{min} unbound plasma concentrations above ~0.9 × EC_{90} (Fig. 4C), as defined in the dNHBE primary cell assay. The 1000 mg/kg BID dose of PF-07321332 maintained a C_{min} unbound plasma concentrations of ~4 × EC_{90} (Fig. 4C). This confirms that PF-07321332 is effective at reducing SARS-CoV-2 MA10 viral load in mouse lungs at concentrations consistent with the observed in vitro antiviral potency and those being targeted clinically.

Disease in this model is manifest by weight loss and pathological changes in the lungs of the infected mice (*33*). Histopathological analysis and immunostaining of lungs from the SARS-CoV-2 MA10–infected mice showed that PF-07321332 limits cellular infiltration (Fig. 4D and fig. S2) and protects lung tissue from damage caused by virus replication (Fig. 4E). Immunohistochemical analysis using a viral nucleocapsid antibody to detect viral antigen levels in the lungs revealed that PF-07321332 inhibits virus replication in a dose-dependent manner (Fig. 4E). Lungs from vehicle-treated, infected mice showed the strongest staining, whereas mock-infected lungs were negative for nucleocapsid staining. Histopathological evaluation of lungs from the vehicle-treated mice demonstrated evidence of increased perivascular inflammation, bronchial or bronchiolar epithelial degeneration or necrosis, bronchial or bronchiolar inflammation, cellular debris in alveolar lumen, and alveolar inflammation and thickening of the alveolar septum compared with PF-07321332–treated mice and mock-infected mice (fig. S2). Most of the infected mice exhibited multifocal pulmonary lesions, which were significantly reduced in PF-07321332–treated mice.

PF-07321332 exhibited moderate plasma clearance (CL_p) in rats and monkeys, with elimination half-lives ($t_{1/2}$) of 5 hours and <1 hour, respectively, after intravenous dosing (table S6). After oral administration to rats, crystalline PF-07321332 (10 mg/kg) demonstrated $F_a \times F_g$ and F values ranging from 65 to 95% and 34 to 50%, respectively, depending on the crystalline form used (Fig. 1 and table S6). Oral administration of PF-07321332 (10 mg/kg) to monkeys led to a relatively poor F of 8.5% ($F_a \times F_g$ = 20%) (table S6), which is attributed to first-pass metabolism along the gastrointestinal tract by cytochrome P450 (CYP) enzymes, consistent with a rapid ($t_{1/2}$ = 20.5 min, CL_{int} = 33.8 μl/min/mg), NADPH-dependent metabolic turnover of PF-07321332 in monkey intestinal microsomes (table S7). PF-07321332 was resistant ($t_{1/2}$ > 240 min, CL_{int} <2.89 μl/min/mg) to CYP-mediated metabolism in intestinal microsomes from rat and human. PF-07321332 (0.3 to 10 μM) exhibited concentration-independent plasma protein binding in rat [mean plasma unbound fraction ($f_{u,p}$) = 0.478], monkey (mean $f_{u,p}$ = 0.434), and human (mean $f_{u,p}$ = 0.310) under equilibrium dialysis conditions (*34*).

Drug-metabolizing enzymes involved in the metabolism of PF-07321332 were also studied.

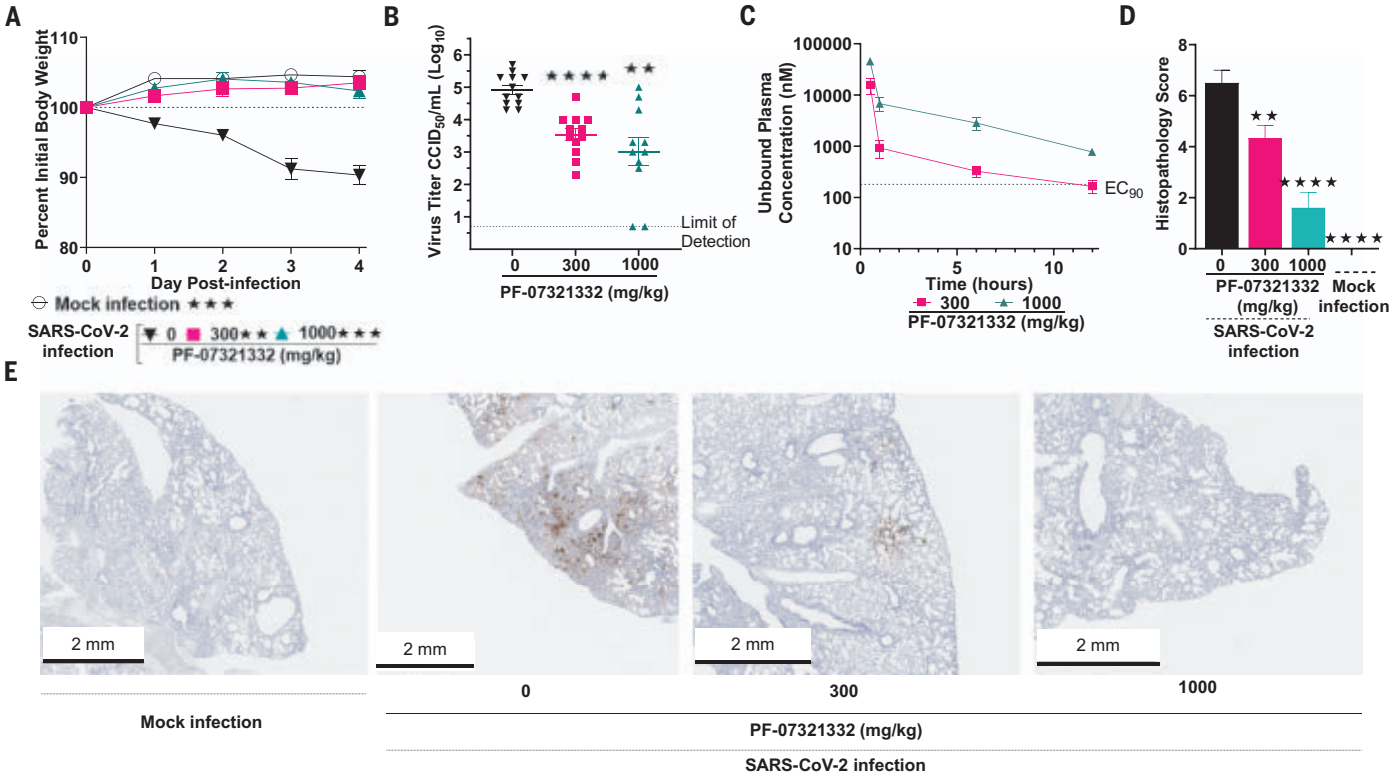


Fig. 4. In vivo efficacy of PF-07321332 against SARS-CoV-2 MA10 infection in mice. Six mice per group were challenged intranasally with $1 \times 10^{5.0}$ 50% $CCID_{50}$ of SARS-CoV-2 MA10. Animals were orally administered 300 or 1000 mg/kg BID PF-07321332 or vehicle (placebo) 4 hours after infection. Animals were euthanized at 4 days postinfection (dpi) and lungs collected for virus titers. Data for (A) to (D) were compiled from two independent studies (n =12 BALB/c mice). (**A**) Weight loss during infection. Mice were weighed daily. (**B**) Lung viral titer at 4 dpi. Lung titers are graphed as mean $\log_{10}$ $CCID_{50}$/ml ± SEM. Dotted line represents the limit of detection for the $CCID_{50}$ assay. (**C**) Twelve-hour PF-07321332 exposure levels of 300 and 1000 mg/kg doses in uninfected, orally treated mice. EC_{90} represented as determined in the day 3 dNHBE primary cell assay (**D**) Histopathology scores on a scale of 0 to 5, where 0 is a normal healthy lung and 5 is severe coalescing areas of necrosis and confluent areas of inflammation. (**E**) SARS-CoV-2 nucleocapsid protein immunohistochemistry. Shown are digital light microscopic scans of mouse lung tissue sections of mock-infected, placebo, 300 mg/kg PF-07321332–, and 1000 mg/kg PF-07321332–treated mice stained with SARS-CoV-2 nucleocapsid antibody. Data are scans from one study. Scale bars, 2 mm. Magnification is 1×.

In NADPH-supplemented HLMs, PF-07321332 demonstrated moderate CL_{int} (24.5 µl/min/mg) (Fig. 1 and table S7), which was significantly inhibited (≥82%) by the selective CYP3A4/5 inhibitor ketoconazole (*28*) (table S8). Moreover, the oxidative metabolic profile of PF-07321332 in HLMs, which includes modifications on the P2 6,6-dimethyl-3-azabicyclo[3.1.0]hexane, the *tert*-butyl group at the P3 position, and the P1 pyrrolidinone ring, was reproduced in incubations of PF-07321332 with recombinant human CYP3A4 (fig. S1). These in vitro studies, which established a predominant role for CYP3A4 in the metabolism of PF-07321332, also presented an opportunity to boost therapeutic concentrations of PF-07321332 in the clinic by co-dosing with the potent CYP3A4 inactivator ritonavir (RTV), which is used as a pharmacokinetic enhancer of several marketed protease inhibitors (e.g., darunavir and lopinavir) that are subject to metabolic clearance through CYP3A4 (*35*, *36*).

PF-07321332 demonstrated a favorable off-target selectivity profile in a broad panel of G protein–coupled receptors, kinases, transporters, and phosphodiesterase enzyme inhibitor screens, and was devoid of activity against the cardiac ion channels $K_v1.1$, $Ca_v1.2$, and $Na_v1.5$ (tables S9 and S10). PF-07321332 was not mutagenic or clastogenic in in vitro genetic toxicity studies and was negative in an in vivo rat micronucleus assay (table S11). Repeat oral dosing of PF-07321332 in 2-week regulatory toxicity studies in monkeys (60 to 600 mg/kg) and rats (40 to 1000 mg/kg) led to dose-dependent increases in both maximal plasma concentrations (C_{max}) and area-under-the-plasma concentration versus time curves (AUCs) (Fig. 5, A and B, and table S12). This resulted in unbound C_{max} and average concentrations (C_{av}) margins of 273-fold and 65-fold in rats, respectively (day 14, 1000 mg/kg), and 510-fold and 245-fold in monkeys, respectively (day 15, 600 mg/kg), over the measured unbound EC_{90} value of PF-07321332 determined in the SARS-CoV-2 day 3 dNHBE cellular assay. PF-07321332 was well tolerated, with no adverse findings in either species; the corresponding no observed adverse effect levels (NOAELs) were the highest doses tested (600 mg/kg/day in monkeys and 1000 mg/kg/day in rats).

The safety, tolerability, and pharmacokinetics of PF-07321332 as a single agent and in combination with RTV are under investigation in a randomized, double-blind, placebo-controlled, single ascending dose study in healthy adult participants (table S13; www.ClinicalTrials.gov identifier: NCT04756531). At each dose tested, four participants were randomized to receive active treatment and two participants received placebo. In the PF-07321332/RTV co-administration dosing paradigm, each subject (active and placebo) received one tablet (100 mg) of RTV at –12 hours, 0 hours, and 12 hours. PF-07321332 was administered as an oral suspension under fasted conditions at 0 hours (minimum fast of ~10 hours before treatment). Preliminary plasma concentration versus time pharmacokinetic profiles achieved from two oral doses, PF-07321332 150 mg alone and PF-07321332 250 mg with RTV, are presented in Fig. 5C. PF-07321332

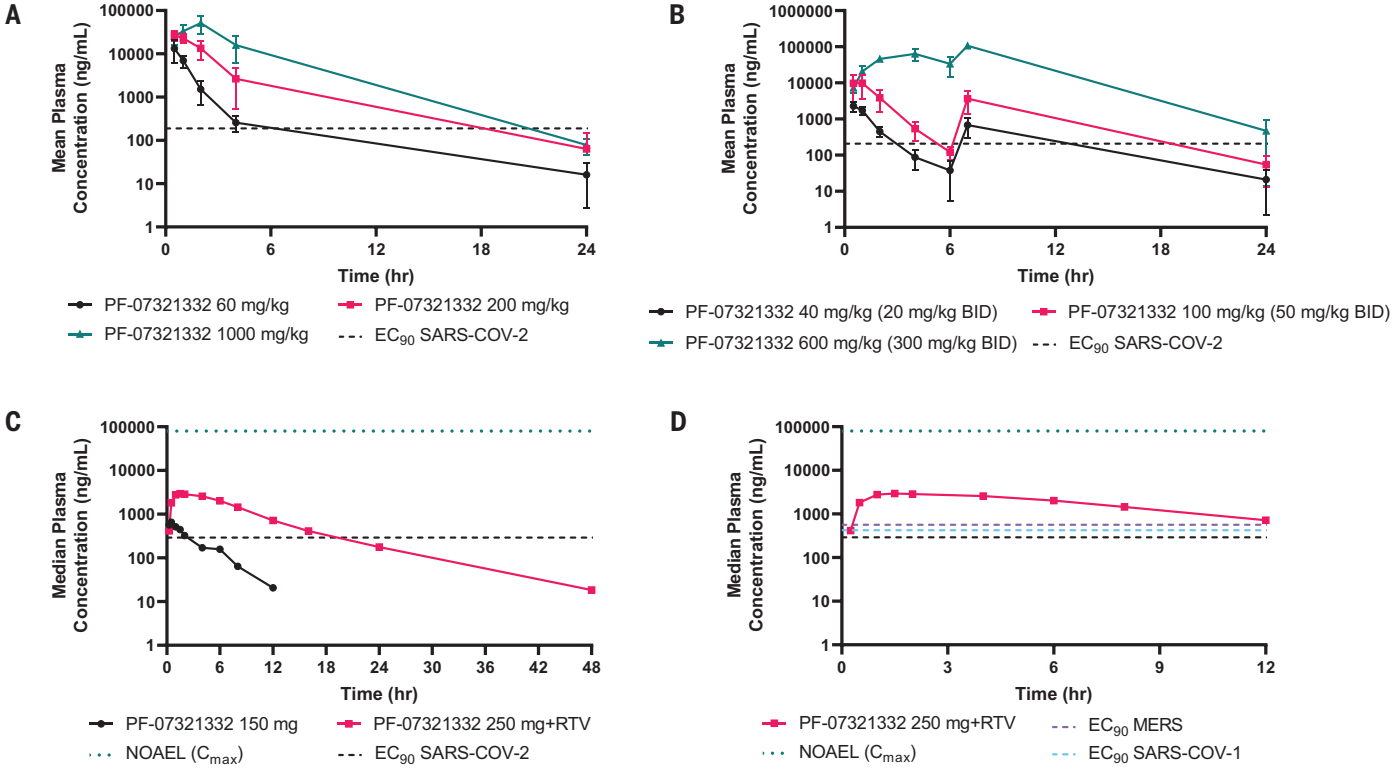


Fig. 5. Preclinical toxicology and healthy adult participant single ascending dose study exposures for PF-07321332. (**A**) Rat oral toxicokinetic exposures (day 14) of once-daily administered PF-07321332 compared with day 3 antiviral EC_{90} values in dNHBE cells. (**B**) Monkey oral toxicokinetic exposures (day 15) of twice-daily administered PF-07321332 compared with day 3 antiviral EC_{90} in dNHBE. (**C**) Human plasma concentrations (total) versus time profile after oral administration (fasted state) of PF-07321332 (150 mg) and PF-07321332 (250 mg) with RTV (100 mg at t = –12 hours, 0 hours, and 12 hours) compared with day 3 antiviral EC_{90} in dNHBE. (**D**) Human antiviral target coverage >$EC_{90,u}$ at 12 hours for SARS-CoV-1, SARS-CoV-2, and MERS-CoV after oral administration of PF-07321332 (250 mg) and ritonavir (100 mg at t = –12 hours, 0 hours, and 12 hours). The in vitro unbound SARS-COV-2 EC_{90} of 181 nM was converted to nanograms per milliliter using a molecular weight of 499.5 g/mol for PF-07321332. Total EC_{90} was calculated by dividing unbound EC_{90} by $f_{u,plasma}$ [rat 0.479, nonhuman primate (NHP) 0.208, and human 0.310]. This resulted in total EC_{90} values of 189, 208, and 292 ng/ml for rat, NHP, and human, respectively. The calculated total human EC_{90} for MERS-CoV and SARS-COV-1 were 566 and 422 ng/ml, respectively. A human NOAEL of 79,700 ng/ml (for C_{max}) was estimated from the rat C_{max} value of 51,500 ng/ml at the NOAEL dose of 1000 mg/kg, normalizing for plasma unbound fraction differences.

was safe and well tolerated and exhibited a significant boost in plasma concentrations when co-administered with RTV. Oral plasma concentrations of PF-07321332 250 mg with RTV were considerably above the SARS-COV-2 antiviral EC_{90} value (total EC_{90} = 292 ng/ml, unbound EC_{90} = 90.5 ng/ml, 181 nM) at 12 hours after dose, thus increasing confidence in achieving robust pan-coronavirus antiviral activity clinically (Fig. 5D). On the basis of these observations and the high safety margins at the NOAEL doses in animals, the efficacy of PF-07321332 in COVID-19 patients will be assessed with a twice-daily dosing paradigm with the potential to increase the dose of PF-07321332 as a single agent and/or co-administered with RTV.

REFERENCES AND NOTES

1. V. M. Corman, D. Muth, D. Niemeyer, C. Drosten, *Adv. Virus Res.* **100**, 163–188 (2018).
2. WHO Solidarity Trial Consortium, *N. Engl. J. Med.* **384**, 497–511 (2021).
3. L. Riva *et al.*, *Nature* **586**, 113–119 (2020).
4. A. Wahl *et al.*, *Nature* **591**, 451–457 (2021).
5. S. S. Good *et al.*, *Antimicrob. Agents Chemother.* **65**, e02479–e02420 (2021).
6. F. Wu *et al.*, *Nature* **579**, 265–269 (2020).
7. P. Zhou *et al.*, *Nature* **579**, 270–273 (2020).
8. T. Pillaiyar, M. Manickam, V. Namasivayam, Y. Hayashi, S. H. Jung, *J. Med. Chem.* **59**, 6595–6628 (2016).
9. Z. Jin *et al.*, *Nature* **582**, 289–293 (2020).
10. I. Schechter, A. Berger, *Biochem. Biophys. Res. Commun.* **27**, 157–162 (1967).
11. W. Rut *et al.*, *Nat. Chem. Biol.* **17**, 222–228 (2021).
12. K. Anand, J. Ziebuhr, P. Wadhwani, J. R. Mesters, R. Hilgenfeld, *Science* **300**, 1763–1767 (2003).
13. L. Zhang *et al.*, *J. Med. Chem.* **63**, 4562–4578 (2020).
14. A. K. Ghosh, H. L. Osswald, G. Prato, *J. Med. Chem.* **59**, 5172–5208 (2016).
15. R. Cannalire, M. L. Barreca, G. Manfroni, V. Cecchetti, *J. Med. Chem.* **59**, 16–41 (2016).
16. J. Qiao *et al.*, *Science* **371**, 1374–1378 (2021).
17. R. L. Hoffman *et al.*, *J. Med. Chem.* **63**, 12725–12747 (2020).
18. L. Zhang *et al.*, *Science* **368**, 409–412 (2020).
19. M. de Vries *et al.*, *J. Virol.* **95**, e01819–e01820 (2021).
20. B. Boras *et al.*, *bioRxiv* 293498 [Preprint] (2020). https://www.biorxiv.org/content/10.1101/2020.09.12.293498v2.
21. L. Di *et al.*, *J. Pharm. Sci.* **100**, 4974–4985 (2011).
22. D. F. Veber *et al.*, *J. Med. Chem.* **45**, 2615–2623 (2002).
23. J. B. Moon, R. S. Coleman, R. P. Hanzlik, *J. Am. Chem. Soc.* **108**, 1350–1351 (2002).
24. C.-P. Chuck *et al.*, *Eur. J. Med. Chem.* **59**, 1–6 (2013).
25. S. Konno *et al.*, *Bioorg. Med. Chem.* **21**, 412–424 (2013).
26. P. S. Dragovich *et al.*, *Bioorg. Med. Chem. Lett.* **10**, 45–48 (2000).
27. M. Kato *et al.*, *Drug Metab. Pharmacokinet.* **18**, 365–372 (2003).
28. R. L. Walsky, R. S. Obach, *Drug Metab. Dispos.* **32**, 647–660 (2004).
29. S. I. Hattori *et al.*, *Nat. Commun.* **12**, 668 (2021).
30. J. Cui, F. Li, Z. L. Shi, *Nat. Rev. Microbiol.* **17**, 181–192 (2019).
31. A. Barilli *et al.*, *Biomedicines* **8**, 127 (2020).
32. M. B. Reddy *et al.*, *Antimicrob. Agents Chemother.* **56**, 3144–3156 (2012).
33. S. R. Leist *et al.*, *Cell* **183**, 1070–1085.e12 (2020).
34. M. J. Banker, T. H. Clark, J. A. Williams, *J. Pharm. Sci.* **92**, 967–974 (2003).
35. K. McKeage, C. M. Perry, S. J. Keam, *Drugs* **69**, 477–503 (2009).
36. R. S. Cvetkovic, K. L. Goa, *Drugs* **63**, 769–802 (2003).

ACKNOWLEDGMENTS

The authors thank the participants of the first-in-human (FIH) study. We would also like to acknowledge the many Pfizer colleagues who have contributed to the COVID-19 oral protease program across a number of disciplines. In particular, we acknowledge S. Sakata, J. Arcari, and J. Zhang for external research resourcing; K. Farley for NMR studies; L. Lanyon for protease panel data; Y. Lian, E. LaChapelle, S. Wright, S. O'Neil, E. Yang, J. Humphrey, and B. Boscoe for compound synthesis; S. Jenkinson for safety pharmacology data; K. Ryan for structural biology support; E. Collins and C. Allais for FIH-enabling active pharmaceutical ingredient supply; F. Clark for bioanalysis; H. Shi for clinical assay support; G. Nucci and A. Bergman for first-in-human study design and clinical pharmacology; F. Hackman for

clinical statistics; S. Toussi for medical monitoring; K. Bartsch for clinical study leadership; S. Pawlak for clinical trial leadership; C. Fredette for clinical trial project management; D. Ding for regulatory documentation support; and M. Dolsten for scientific discussion and advice. This research used resources of the Advanced Photon Source, a US Department of Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by Argonne National Laboratory under contract no. DE-AC02-06CH11357. Extraordinary facility operations were supported in part by the DOE Office of Science through the National Virtual Biotechnology Laboratory, a consortium of DOE national laboratories focused on the response to COVID-19, with funding provided by the Coronavirus CARES Act. **Funding:** This study was sponsored by Pfizer, Inc. **Author contributions:** Conceptualization: C.M.N.A., A.S.A., D.R.O., M.P.; Formal analysis: L.D., B.B., S.E.G., Q.Y., D.K.R., J.J.N., A.D., R.S.O., R.S.P.S.; Investigation: M.A., S.B., J.C.L., J.L., K.O., L.W., R.F., K.S.G., W.L., R.S.O., D.K.R., S.A.G., L.A., S.N., H.E.; Methodology: J.B.T., Q.Y., M.F.S., P.R.V., M.R.R., M.P., D.R.O., N.C.P., A.C., E.P.K., K.J.C., R.S.P.S.; Project administration: D.R.O., M.R.R., J.B.T., A.S.K., C.M.S., L.U., J.G.S., B.L.H., Y.Z., S.W.M., R.D.C., N.C.P.; Resources: A.D., J.J.N., K.J.C.; Supervision: M.R.R., C.E.S., A.E.S., B.L.H., P.R.V., R.D.C.; Visualization: S.W.M., D.K.R., S.N., B.B., L.A., H.E.; Writing – original draft: D.R.O., A.S.K., M.F.S., C.M.S., R.D.C., M.R.R., J.G.S., C.M.S., Y.Z.; Writing – review and editing: D.R.O., A.S.K., M.F.S., A.S.A., C.M.N.A. **Competing interests:** D.R.O., C.M.N.A., A.S.A., L.A., M.A., S.B., B.B., R.D.C., A.C., K.J.C., A.D., L.D., H.E., R.F., K.S.G., S.E.G., E.P.K., A.S.K., J.C.L., J.L., W.L., S.N., R.S.O., K.O., N.C.P., D.K.R., M.R.R., M.F.S., J.G.S., R.P.S., C.M.S., A.S., J.B.T., L.U., P.R.V., L.W., Q.Y., and Y.Z. are employees of Pfizer and some of the authors are shareholders in Pfizer Inc. S.W.M., J.J.N., and M.P. were employees of Pfizer Inc. during part of this study. **Data and materials availability:** All nonclinical data are available in the main text or the supplementary materials. X-ray coordinates and structure factors are deposited at the RCSB Protein Data Bank under accession codes 7RFR, 7RFU, 7RFS, and 7RFW. Individualized Fig. 5 data and the statistical analysis plan are available on www.Figshare.com with the following DOI: 10.25454/pfizer.figshare.16699669. Pfizer compound use requests are processed through the Pure Compound Grants program (see https://www.cybergrants.com/pfizer/Research). Pfizer Inc. has applied for patent applications covering PF-07321332 as well as related compounds. A phase 1 clinical trial has been registered at www.ClinicalTrials.gov under identifier NCT04756531. 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.org/doi/10.1126/science.abl4784
Materials and Methods
Figs. S1 and S2
Tables S1 to S13
References (*37*–*70*)
MDAR Reproducibility Checklist

15 July 2021; accepted 28 October 2021
Published online 2 November 2021
10.1126/science.abl4784

ELECTROCHEMISTRY

Enhancement of lithium-mediated ammonia synthesis by addition of oxygen

Katja Li[1]†, Suzanne Z. Andersen[1]†, Michael J. Statt[2]†, Mattia Saccoccio[1], Vanessa J. Bukas[1]‡, Kevin Krempl[1], Rokas Sažinas[1], Jakob B. Pedersen[1], Vahid Shadravan[1], Yuanyuan Zhou[1], Debasish Chakraborty[1], Jakob Kibsgaard[1], Peter C. K. Vesborg[1], Jens K. Nørskov[1]*, Ib Chorkendorff[1]*

Owing to the worrying increase in carbon dioxide concentrations in the atmosphere, there is a need to electrify fossil-fuel–powered chemical processes such as the Haber-Bosch ammonia synthesis. Lithium-mediated electrochemical nitrogen reduction has shown preliminary promise but still lacks sufficient faradaic efficiency and ammonia formation rate to be industrially relevant. Here, we show that oxygen, previously believed to hinder the reaction, actually greatly improves the faradaic efficiency and stability of the lithium-mediated nitrogen reduction when added to the reaction atmosphere in small amounts. With this counterintuitive discovery, we reach record high faradaic efficiencies of up to 78.0 ± 1.3% at 0.6 to 0.8 mole % oxygen in 20 bar of nitrogen. Experimental x-ray analysis and theoretical microkinetic modeling shed light on the underlying mechanism.

Ammonia (NH_3) is one of the most abundantly manufactured chemicals worldwide, with a yearly production of over 182 million tonnes (*1*). Its main use is as a synthetic fertilizer (~80%) and as the source of all activated nitrogen in the chemical industry, but it has recently also been considered as a carbon-free energy carrier (*2*–*4*). Currently, ammonia is produced from nitrogen and hydrogen through the thermally catalyzed Haber-Bosch process, which operates under harsh conditions (350° to 450°C, 100 to 200 bar), requiring large centralized plants and high capital investment (*5*, *6*). To satisfy the commercial demands, about ~1% of global energy consumption is used in the process (*7*). Additionally, the Haber-Bosch process is responsible for about 1.4% of the annual CO_2 emissions, as the supplied H_2 originates from steam reforming of methane (*8*–*10*). An alternative, environmentally sustainable way to produce ammonia is through an electrochemical pathway, with the electrical energy provided from renewable sources such as wind or solar power. Recently, efforts toward electrochemical synthesis of ammonia have increased substantially (*11*–*14*); however, the field has been hampered by various issues. One major concern in the literature is related to contamination of the input gases, chemicals, and catalysts by ammonia and other labile nitrogen compounds (*15*, *16*), which may result in an erroneously high reported faradaic efficiency (FE). Several protocols (*16*–*18*) have been published on proper contaminant identification and rigorous experimentation, and some of the erroneous reports are being corrected or withdrawn as scientists retest and reevaluate methods and systems (*19*, *20*). In a recent paper, Choi *et al.* investigated over 130 publications on electrochemical ammonia synthesis, concluding it highly likely that none of the aqueous systems produce ammonia and that the most reliable method presently is lithium-mediated nitrogen reduction (LiNR) in nonaqueous electrolytes (*16*). A similar lithium-mediated method was first published by Fichter *et al.* in 1930 (*21*) and later investigated with a near-aprotic solvent by Tsuneto *et al.* in the 1990s (*22*, *23*). Currently, the Tsuneto-based system has been revisited by several groups (*11*, *12*, *17*, *18*, *24*–*28*); however, the exact mechanism is still not fully understood. It is generally accepted that the LiNR takes place in three steps (*25*), the first one being the electrochemical reduction of Li^+ ions in the electrolyte to metallic Li, which is a very reactive material. This freshly plated Li is believed to then dissociate N_2, and the N at the surface is finally reduced in a series of electron and proton transfers to form NH_3 by using a suitable proton source, like ethanol (EtOH). An important component of the LiNR system is the solid electrolyte interface (SEI) that forms from decomposition products of an organic electrolyte during Li deposition on the cathode. The SEI provides a porous passivation layer over the electrode surface that is electronically insulating but ionically conducting. Its exact composition and mechanistic role in the LiNR process are still unclear and difficult to determine, because it is strongly dependent on experimental conditions and sensitive to air exposure. Nevertheless, the process reliably forms ammonia from N_2 and a proton source at ambient conditions, typically achieving a FE of around 5 to 20% (*17*, *25*, *29*), with a recent breakthrough in FE of up to 69% by Suryanto *et al.* (*11*) under 20-bar N_2. Lazouski *et al.* implemented this process using a gas diffusion electrode (GDE) setup with reported efficiencies of 30%, but the

[1]Department of Physics, Technical University of Denmark, Kongens Lyngby, Denmark. [2]SUNCAT Center for Interface Science and Catalysis, Department of Chemical Engineering, Stanford University, Stanford, CA, USA.
***Corresponding author. Email: jkno@dtu.dk (J.K.N.); ibchork@fysik.dtu.dk (I.C.)**
†These authors contributed equally to this work.
‡Present address: Fritz-Haber-Institut der Max-Planck-Gesellschaft, 14195 Berlin, Germany.

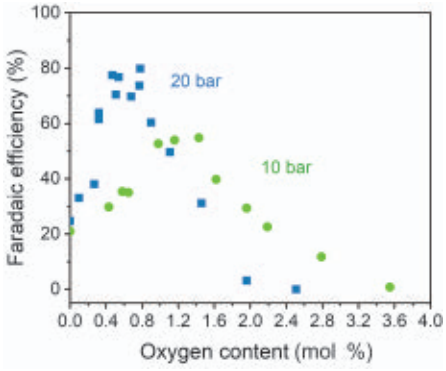


Fig. 1. Influence of oxygen content on the FE of the Li-mediated ammonia synthesis. Reactions were conducted at 10 (green) and 20 (blue) bar, where the added oxygen content inside the autoclave was determined by mass spectrometry. The reported pressure is the total pressure of N_2 and O_2 combined. All experiments were operated at a constant current of −4 mA/cm^2 until the system overloaded or reached 50 C (116 min to pass 50 C), unless stated otherwise.

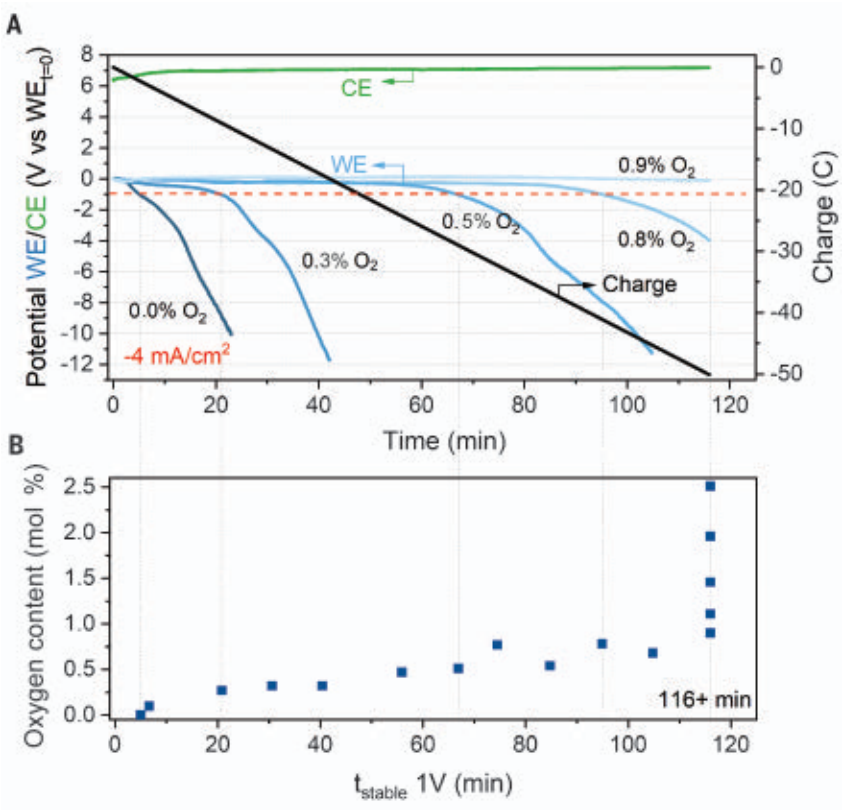


Fig. 2. Influence of oxygen content on the stability of the Li-mediated ammonia synthesis. (**A**) Chronopotentiometry with −4 mA/cm^2 applied at 20 bar N_2 with variable O_2 content. The potentials are normalized to the starting WE potential. A maximum of 50 C charge (black trace) is passed; however, some measurements overloaded before 50 C was reached, as seen from the sudden decrease in WE potential (blue traces) for low O_2 concentrations (all shown % O_2 are mol % O_2). A representative CE potential (green trace) is shown, which is stable throughout the measurement, even for experiments in which the WE overloaded. The dotted red line indicates the decrease of WE potential by 1 V. (**B**) The time until the WE potential decreased by 1 V (t_{stable}), as a function of O_2 content. All stable experiments were included in 116+ min.

system was only stable for a few minutes and exhibited high cell potentials (*12*). Instability of this process was already described by Tsuneto *et al.* (*22*), and recently mitigated by Andersen *et al.* with a potential cycling strategy (*24*), which enabled stability over several days and increased the FE at 10-bar N_2 from 20% without cycling to 37% with cycling at an energy efficiency (EE) of 7%. However, the REFUEL program of the US Department of Energy set a target of 90% FE at 300 mA/cm^2 and an EE of 60% (*30*). The current state of the lithium-mediated process is clearly far from this target. Especially the EE is currently a major problem in the LiNR, because Li plating requires largely negative potentials (−3.04 V versus SHE). The overpotential losses of the LiNR are portrayed and discussed in fig. S1. If all overpotentials are minimized and hydrogen oxidation reaction (HOR) is utilized at the counter electrode (CE), the resulting EE would be 26%, assuming 80% FE.

Here we show that adding small amounts of O_2 to the feed gas has a positive effect on both the FE and the stability of the system. A FE of up to 78.0 ± 1.3% at 20-bar N_2 can be achieved by adding 0.5 to 0.8 mol % O_2, resulting in an EE of 11.7 ± 0.5% (calculation in supplementary materials). This EE calculation accounts for neither EtOH as sacrificial proton donor nor for the energy of pressurizing the system. We used this framework as a basis for comparison to results from literature (*24*, *25*). Using hydrogen oxidation at the CE in a GDE-type system has previously been shown to prevent solvent oxidation (*12*), and the use of a phosphonium proton shuttle has been experimentally verified to successfully shuttle the newly created protons to the working electrode while the phosphonium ylide becomes reprotonated at the anode (*11*), thereby circumventing the sacrificial ethanol issue. Our focus in this study lies on the finding that oxygen increases the FE and not on the sacrificial proton donor issue, but we expect that the previously reported solutions will also apply to our system. The positive effect of small amounts of oxygen is counterintuitive; one would expect oxygen to contaminate the active phase and result in loss of efficiency because of the oxygen reduction reaction (ORR). Indeed, early experiments by Tsuneto *et al.* showed that higher oxygen content in the gas feed (synthetic air with 20% O_2) lowered the FE considerably from ~50 to 0.1% at 50 bar (*23*). We investigated the origin of the oxygen promotion using a microkinetic model, suggesting that the higher FE is related to slower Li^+ diffusion through the SEI layer formed in the presence of O_2 as observed by Wang *et al.* in studies of Li–air battery materials (*31*). We further tested this hypothesis by systematically investigating the effect of O_2 on the SEI layer through x-ray diffraction (XRD) and x-ray photoelectron spectroscopy (XPS).

Effect of oxygen on Faradaic efficiency

Systematic experiments with varying amounts of O_2 added at 10- and 20-bar total gas pressure were conducted. All experiments were performed in an in-house–fabricated autoclave based on the design by Wiberg *et al.* (*32*) and carried out in a glass cell containing ~30 ml of electrolyte (fig. S2). The gases used for the experiments were cleaned with purifiers (NuPure) to levels as low as parts per trillion by volume of nitrogen-containing impurities, and the flow was controlled by mass flow controllers (Brooks) to ensure proper gas mixing. This was important, as improper gas mixing affects the electrochemical reaction (figs. S3 and S4). The oxygen content was determined by a mass spectrometer (Pfeiffer, omnistar GSD 320) attached to the autoclave with a 1-μm orifice just above the electrochemical cell. A representative mass spectrum is shown in

fig. S5. The electrolyte was 0.3 M $LiClO_4$ in tetrahyrofuran (THF) with 1.0 vol % EtOH, unless otherwise specified. The working electrode (WE) was a Mo foil of 1.8 cm_{geo}^2 area, the CE was a 1.8 cm_{geo}^2 Pt mesh, and the pseudo-reference electrode (RE) was a Pt wire. The potential scale was determined by calibrating the Pt pseudo-RE to the well-known Fc/Fc^+ couple (fig. S6). A constant-current density of -4 mA/cm^2 was applied, until the system either overloaded because of an increased total cell potential beyond the capacity of the potentiostat, or the experiment reached 50 C of total passed charge. A representative cyclic voltammetry is shown in fig. S7. Cycling experiments according to Andersen *et al.* (*24*) were also conducted, to show the beneficial coupling of the improved stability due to cycling with the increased FE from the addition of O_2. The FE was determined at the end of the experiment and therefore represents accumulated FE. A more detailed description of the experimental procedure and all experimental results can be found in the supplementary materials.

Figure 1 shows that small amounts of O_2 markedly increase the FE toward ammonia production at both 10 and 20 bar, showing peak behavior from 0.5 to 0.8 mol % O_2 in N_2 at 20 bar and 1.2 to 1.6 mol % at 10 bar. The optimum oxygen content at 10 bar is higher than at 20 bar by almost exactly a factor of 2, suggesting that the oxygen partial pressure, rather than the molar ratio of O_2 in N_2, is responsible for the increase in FE (fig. S8). The efficiency drops rapidly at higher amounts of O_2, in accordance with the much higher oxygen-content experiments (20% O_2, 50 bar) of Tsuneto *et al.* (*23*). An increase in the H_2O content of the electrolyte postelectrochemistry is observed with increasing O_2 content (fig. S9), suggesting that ORR is responsible for the decrease in FE. The decrease in the FE at higher O_2 content could also partly be due to the formation of passivating Li_2O species on the surface of the WE as seen later by XRD. As expected, the EE of the system follows a trend similar to that of the FE with peaks at 8.0 ± 1.8% and 11.7 ± 0.5% at optimal O_2 content under 10 bar and 20 bar, respectively (figs. S10 and S11).

Effect of oxygen on stability

In addition to the notable increase in the FE of the system because of added O_2, an increase in the apparent stability of the total cell potential was also observed. However, the electrochemistry will overload given enough time, and the only way to ensure stability in this system is through potential cycling (*24*). Figure 2A shows representative electrochemistry data of chronopotentiometry measurements at 20 bar, where it is clear that the WE potential remains stable for longer times with increasing O_2 content. The CE remains stable, even for experiments in which the WE is unstable. This excludes the possibility of oxygen having an appreciable influence on the CE. To compare the stabilities, we plotted the data against the starting WE potential of each individual experiment. The total time it takes for the WE potential to decrease by 1 V was taken as a measure of stability (t_{stable}). Because all experiments exceeding 0.8 mol % O_2 exhibited a WE potential that did not decrease by more than 1 V, t_{stable} was set to 116+ min for these experiments. t_{stable} is plotted as a function of the O_2 content in Fig. 2B, and a linear trend of increasing stability by increasing O_2 is observed. Stability remained high even for larger O_2 content when FE diminished. This relation can be visualized in fig. S10A and suggests that LiNR stability is qualitatively related to O_2 content but not to the FE. Cycling of the potential to prevent build-up of passivating lithium species is a previously proven method to achieve long-term stability (days) of the potential in the LiNR system (*24*). This method was applied for an experiment with optimal O_2 content (20 bar N_2, 0.67 mol % O_2) and one with suboptimal O_2 content (20 bar N_2, 0.26 mol % O_2), which would otherwise be unstable under constant Li deposition. We found that cycling the potential in the presence of O_2 retains the higher FE (fig. S9) while also stabilizing the WE during the 0.26 mol % O_2 experiment until (at least) 50 C of charge has been passed (figs. S12 and S13). In addition, the water content of these cycling experiment increased substantially compared to the constant-current experiments (table S1), owing to the constant ORR. However, because the FE seems to be unaffected by this result, it is believed that the increase in water on this scale is not detrimental to the system and that by applying cycling on the optimum oxygen content of 0.6 to 0.8 mol % at 20 bar, the reaction could be kept stable for longer (10 hours), as was shown previously (*24*).

Mechanistic role of O_2 in LiNR

We recently demonstrated that in the absence of oxygen, the LiNR system and its FE toward ammonia are largely controlled by mass transport limitations of the reactant species diffusing from the bulk electrolyte to the catalyst surface (*24*). By developing a kinetic model, we showed that the experimental FE is a simple function of the relative diffusion rates between incoming lithium ions (r_{Li}), protons (r_H), and nitrogen (r_{N_2}). Here, we assume a steady-state FE, although there might be dynamical changes especially initially where the SEI layer is formed. Our model considered a series of elementary steps and relied on the basic, yet reasonable, assumption that all diffusion steps are considerably slower than subsequent reaction steps taking place at the surface, i.e., they are overall rate-limiting. This assumption results from recognizing that (i) electrochemical surface steps are most likely very fast and driven irreversibly forward because of the extreme reducing potentials required to plate Li (*33*); (ii) the chemical (dissociative) adsorption of N_2 on metallic Li has a negligible kinetic barrier (*34*, *35*); and (iii) the SEI layer that

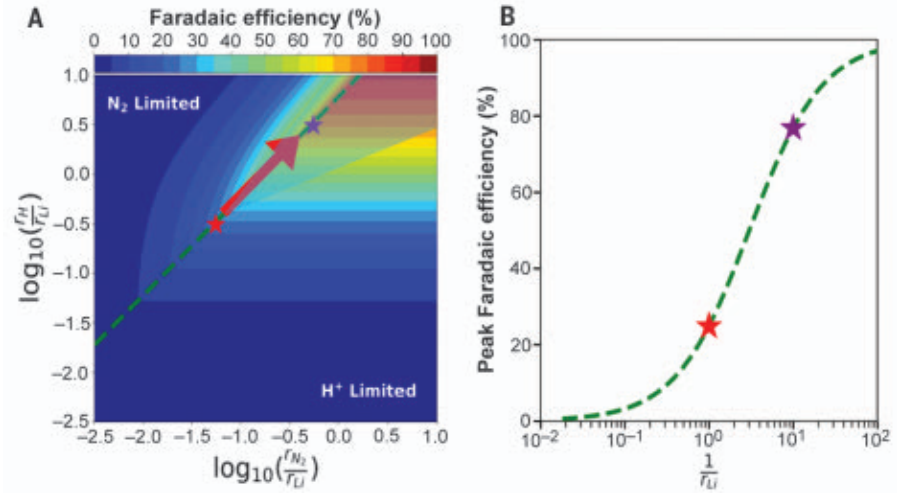


Fig. 3. Graphic illustration of the microkinetic model for Li-mediated ammonia synthesis. (**A**) Heatmap of the predicted FE as a function of the ratio of nitrogen to lithium (*x* axis) and proton to lithium (*y* axis) diffusion rates. The red star indicates the expected location of the 10-bar experiments without O_2 in the system. The purple star indicates the improvement in FE if r_{Li} were selectively lowered by an order of magnitude. The cone represents the uncertainty of the location of the purple star. (**B**) A one-dimensional plot of FE cut along the optimal r_{N_2}/r_H ratio (marked by the green dashed line in Fig. 3A) that clearly shows the marked FE improvement that can result from lowering r_{Li}. The *x* axis is normalized so that the red star corresponds to $1/r_{Li} = 1$.

forms during Li deposition in THF most likely leads to much slower diffusion as compared to that in aqueous media (*36*). Alongside rate-limiting diffusion, we further consider mass balance of the incoming protons toward either H_2 or NH_3 and thus distinguish between two possible mass-transfer limitations for the nitrogen reduction reaction (NRR): a limitation in nitrogen (N_{lim}), or a limitation in protons (H_{lim}). Within these two regimes, the steady-state rate of NH_3 production and resulting FE can be formulated in terms of only the three elementary (r_{Li}, r_H, r_{N_2}) diffusion rates. The heatmap of Fig. 3A shows how the predicted FE changes as a function of the relative-to-Li rates of incoming N_2 (x axis) and H^+ (y axis). A dashed green line marks the ideal r_{N_2} to r_H ratio that leads to the highest FE at the boundary between nitrogen- and proton-limited regimes.

When adding oxygen to the gas feed, it is intuitive to expect that the NRR will be suppressed because of additional competition from the ORR. The experimental improvement in FE at low O_2 content, however, is a counterintuitive result that is much harder to understand. As a first step toward this understanding, we extend our original kinetic model to incorporate the oxygen electrochemistry and examine its effect on the resulting FE. This involves not only the inclusion of the ORR that consumes protons and electrons at the catalyst surface, but also the probability that newly formed H_2O can itself act as a proton donor to further drive the catalysis. Derivation of the enhanced model can be found in the supplementary materials; we emphasize that none of the situations that we considered can account for an oxygen-induced improvement in FE as found experimentally. Indeed, we predict a (likely small) decrease in FE at low O_2 content because of the ORR competing as a parasitic side reaction, and a substantial drop to ≈0 when the more reactive oxygen species dominate all surface sites over nitrogen species at higher O_2 contents. This result suggests that O_2 must have another way of influencing the LiNR system.

What can increase the FE is a decrease in the Li diffusion rate (Fig. 3A). This is simply because Li electrodeposition also competes with the NRR for electrons. Ideally, one would want Li diffusion or deposition to be slow enough that it consumes a minimal number of electrons while still providing a full monolayer of clean, freshly plated Li that is reactive enough to dissociate nitrogen. Delayed Li diffusion as a result of oxygen-induced changes in the SEI layer has, indeed, been observed. Wang *et al.* recently showed that O_2 can specifically influence the SEI composition in nonaqueous lithium–air batteries, leading to increased SEI homogeneity as well as diminished Li^+ diffusivity (*31*). Such an effect on Li^+ diffusivity should have a substantial

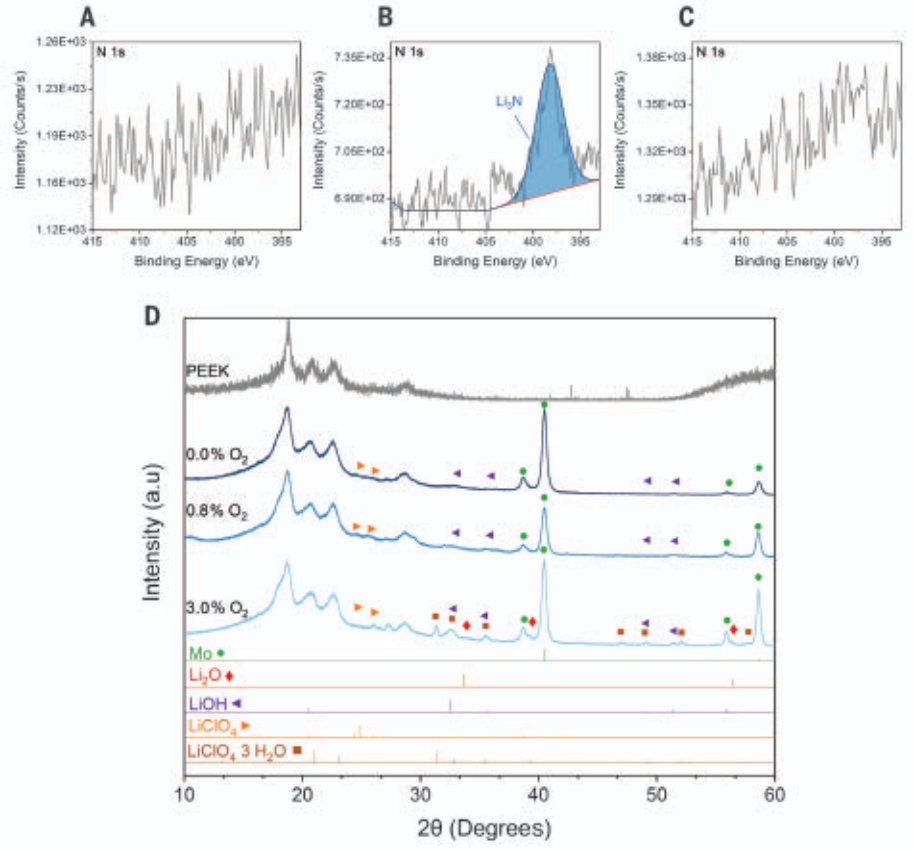


Fig. 4. XPS and XRD data of the working electrode after electrochemistry without exposure to air. Scans of three different WE foils postelectrochemistry with a constant −4 mA/cm^2 applied under 20-bar N_2 with varying amounts of O_2 present in the atmosphere. Foils were transferred without air exposure to the respective systems. (**A** to **C**) XPS spectra of N 1s scanning window for samples with 0.0 mol % (A), 0.8 mol % (B), and 3.0 mol % O_2 (C) in the reaction atmosphere. More high-resolution scans and full survey spectra are included in figs. S17 and S18. The N 1s peak at 0.8 mol % O_2 was seen in two independent measurements. (**D**) XRD diffractograms of samples with 0.0, 0.8, and 3.0 mol % O_2 in atmosphere and pattern of the PEEK dome transfer system, which creates reflections in the 15° to 30° region. Reference patterns for Mo (ICSD: 998-005-2267), Li_2O (ICSD: 98-017-3206), LiOH (ICSD: 98-002-6892), $LiClO_4$ (ICSD: 98-016-5579), $LiClO_4$ $3H_2O$ (ICSD: 98-003-2534), and Li_3N (ICSD: 98-003-4779) are included.

impact on the FE toward forming ammonia. If the rates of nitrogen and proton diffusion, r_{N_2} and r_H, remain unaffected by the oxygen content, a decrease in the Li^+ diffusion rate will move the operating point along the dashed diagonal in Fig. 3A, leading to a substantial increase in the FE. In reality, r_{N_2} and r_H may also change as a result of the oxygen-induced modification of the SEI layer, but as long as the decrease is less than that for r_{Li}, a net increase in the FE is expected, as indicated by the conical shaded area in Fig. 3A.

The improvement in FE, also shown in Fig. 3B, is much larger than what can be expected from moving either horizontally (by, e.g., changing N_2 pressure) or vertically (by, e.g., changing proton activity) across the plot. According to this picture, an optimal O_2 content should follow as a trade-off between FE increasing at low values because of restricted Li^+ diffusivity and decreasing at high values when the ORR starts dominating over the NRR. Furthermore, both the optimal O_2 content and resulting maximum FE should correlate directly with N_2 pressure when the NRR is in the nitrogen-limited regime. As seen from Fig. 3A, increasing N_2 pressure (i.e., moving horizontally to the right on the heatmap) increases FE until the NRR is pushed into the proton-limited regime. The optimal O_2 content is defined by the competition between NRR and ORR and, while increasing N_2 pressure in the nitrogen-limited regime, the maximum FE should occur at same absolute O_2 partial pressure or a proportionally lower O_2 content, exactly as observed experimentally (Fig. 1).

Compositional changes in the SEI layer in the presence of O_2 are also consistent with the observed improvement in stability. For Li-ion batteries, the addition of O_2 has been shown to form a more homogeneous and uniform SEI layer, leading to increased cyclability (*31*). A more homogeneous layer could lead to more uniform Li plating in the LiNR system,

thereby keeping the total cell potential stable for longer. Supporting this hypothesis, we visually observed a change in the deposit on the WE from experiments without O_2 (0.0 mol % O_2), with added O_2 achieving optimum FE (0.5 to 0.8 mol % O_2), and with high O_2 concentrations yielding little to no ammonia (>2.6 mol % O_2). The Mo foil at optimum O_2 content showed a thick deposit on both sides, whereas in the other cases, only a thin deposit on the side facing the CE was observed (fig. S14). This again proves that O_2 affects the WE rather than the CE.

In conclusion, combining our model with a diminished Li^+ diffusivity through the SEI layer in the presence of oxygen explains our experimental observations and suggests that small amounts of O_2 can act as a beneficial SEI additive or modifier to greatly improve both the system's apparent stability and FE toward ammonia.

SEI layer investigation

The key conclusion of the model is that the high FE is a result of a SEI layer modified by the addition of oxygen. To determine the possible compositional effect of O_2 on the SEI layer, we carried out three representative electrochemical experiments at 20 bar with 0.0, 0.8, and 3.0 mol % O_2 in an autoclave kept inside an Ar glove box (fig. S2C). For XPS, the WE foils were loaded into a home-built vacuum transfer system, rapidly evacuated to pressures below 2×10^{-5} mbar, and then loaded into the XPS chamber having a base pressure below 9×10^{-10} mbar (fig. S15). For grazing incidence XRD, the samples were transported in an air-tight polyetheretherketone (PEEK) dome (fig. S16). Hence, both systems prevent air exposure, as the samples are only exposed to the glove box inert Ar atmosphere before loading into the respective transfer systems.

As seen from XPS in Fig. 4, A to C, and fig. S17, the most notable difference between the samples with varying O_2 content is for the sample with 0.8 mol % O_2, wherein a peak appears in the N 1s scanning window. On the basis of the proposed reaction mechanism, which suggests Li_3N as a precursor for ammonia (*25*), we attribute that peak to Li_3N. The peak position also fits well with previously reported spectra of Li_3N (*37*, *38*). The XRD pattern for Li_3N is not observed in Fig. 4 D, because the formed Li_3N is expected to be mainly a surface phase and would also react rapidly with protons to form ammonia. The 0 mol % O_2 sample shows many XPS peaks for Cl compounds, which are not observed in the higher O_2 samples (fig. S17), indicating that these additional species could cause the premature overloading of the system. The XRD diffractograms for the samples appear to increase in complexity with increasing O_2 content, as more compounds appear. For all three samples, the Mo foil is clearly visible, as well as LiOH and $LiClO_4$. The sample with 3.0 mol % O_2 additionally shows hydrated $LiClO_4$, as an increase in O_2 leads to an increase in the ORR (fig. S9 and tables S1 and S2). Furthermore, Li_2O is also faintly visible on this sample, as would be expected on a sample with increased O_2 exposure.

The counterintuitive results shown in this study focus attention on the effect of controlling the availability of competing H^+, N_2, and Li^+ ions at the catalyst surface. Our model suggests that oxygen can slow Li^+ diffusion through the SEI layer while keeping the optimal relationship between available H^+ and N_2 at the surface. The modification of the SEI layer through O_2 is well supported by Wang *et al.* (*31*) and the experimental data presented here. O_2 is shown to have an effect on the deposition behavior (fig. S14), as well as on the stability (Fig. 2). Because it was discussed previously that the increase in stability is caused by more uniform Li plating (*24*), both observations are evidence for O_2 influencing the deposition of Li through modification of the SEI layer. Further optimization of the LiNR system in future studies may direct design of an artificial SEI layer with improved transport-controlling properties, preferably even without using Li as one of the constituents to decrease the overall required cell potential.

We propose that the results presented here put into perspective all previous LiNR publications that did not conduct their experiments entirely in a glove box or with a fully controlled atmosphere. Some experiments could have small amounts of O_2 in the system, which would lead to higher efficiencies but perhaps also to large interexperimental variability (poor reproducibility). Because even very small amounts of O_2 can have such a critical effect, we strongly recommend that the atmospheric composition of the experiment be determined and stated in all future studies. Our findings should also lead to a substantial advantage in the scale-up of the process, because the LiNR will not require ultrapure nitrogen, unlike the Haber-Bosch process, in which even the smallest contamination by oxygen is detrimental to the catalyst (*39*). The separation of N_2 from air is industrially achieved by cryogenic separation in large facilities and is therefore prohibitively expensive for a decentralized system (*40*, *41*).

REFERENCES AND NOTES

1. United States Geological Survey (USGS), *Mineral Commodity Summaries 2020* (2020).
2. D. R. MacFarlane *et al.*, *Adv. Mater.* **32**, e1904804 (2020).
3. D. R. MacFarlane *et al.*, *Joule* **4**, 1186–1205 (2020).
4. C. H. Christensen, T. Johannessen, R. Z. Sørensen, J. K. Nørskov, *Catal. Today* **111**, 140–144 (2006).
5. F. Haber, R. Le Rossignol, *Z. Elektrochem. Angew. Phys. Chem.* **19**, 53–72 (1913).
6. J. W. Erisman, M. A. Sutton, J. Galloway, Z. Klimont, W. Winiwarter, *Nat. Geosci.* **1**, 636–639 (2008).
7. C. J. M. van der Ham, M. T. M. Koper, D. G. H. Hetterscheid, *Chem. Soc. Rev.* **43**, 5183–5191 (2014).
8. M. Capdevila-Cortada, *Nat. Catal.* **2**, 1055 (2019).
9. I. McPherson, J. Zhang, *Joule* **4**, 12–14 (2020).
10. S. T. Wismann *et al.*, *Science* **364**, 756–759 (2019).
11. B. H. R. Suryanto *et al.*, *Science* **372**, 1187–1191 (2021).
12. N. Lazouski, M. Chung, K. Williams, M. L. Gala, K. Manthiram, *Nat. Catal.* **3**, 463–469 (2020).
13. B. Yang, W. Ding, H. Zhang, S. Zhang, *Energy Environ. Sci.* **14**, 672–687 (2021).
14. H. Shen *et al.*, *Chem* **7**, 1708–1754 (2021).
15. J. Kibsgaard, J. K. Nørskov, I. Chorkendorff, *ACS Energy Lett.* **4**, 2986–2988 (2019).
16. J. Choi *et al.*, *Nat. Commun.* **11**, 5546 (2020).
17. S. Z. Andersen *et al.*, *Nature* **570**, 504–508 (2019).
18. H. Iriawan *et al.*, *Nat. Rev. Methods Primers* **1**, 56 (2021).
19. W. Yu *et al.*, *Sustain. Energy Fuels* **4**, 5080–5087 (2020).
20. S. Licht *et al.*, *Science* **369**, 780 (2020).
21. F. Fichter, P. Girard, H. Erlenmeyer, *Helv. Chim. Acta* **13**, 1228–1236 (1930).
22. A. Tsuneto, A. Kudo, T. Sakata, *Chem. Lett.* **22**, 851–854 (1993).
23. A. Tsuneto, A. Kudo, T. Sakata, *J. Electroanal. Chem. (Lausanne)* **367**, 183–188 (1994).
24. S. Z. Andersen *et al.*, *Energy Environ. Sci.* **13**, 4291–4300 (2020).
25. N. Lazouski, Z. J. Schiffer, K. Williams, K. Manthiram, *Joule* **3**, 1127–1139 (2019).
26. T. M. Pappenfus, K. Lee, L. M. Thoma, C. R. Dukart, *ECS Trans.* **16**, 89–93 (2009).
27. L.-F. Gao *et al.*, *Angew. Chem. Int. Ed.* **60**, 5257–5261 (2021).
28. R. Sažinas *et al.*, *RSC Advances* **11**, 31487–31498 (2021).
29. J. A. Schwalbe *et al.*, *ChemElectroChem* **7**, 1542 (2020).
30. G. Soloveichik, in *2019 AIChE Annual Meeting* (AIChE, 2019). https://arpa-e.energy.gov/technologies/programs/refuel
31. E. Wang, S. Dey, T. Liu, S. Menkin, C. P. Grey, *ACS Energy Lett.* **5**, 1088–1094 (2020).
32. G. K. H. Wiberg, M. J. Fleige, M. Arenz, *Rev. Sci. Instrum.* **85**, 085105 (2014).
33. A. R. Singh *et al.*, *ACS Catal.* **9**, 8316–8324 (2019).
34. J. M. McEnaney *et al.*, *Energy Environ. Sci.* **10**, 1621–1630 (2017).
35. T. Ludwig, A. R. Singh, J. K. Nørskov, *J. Phys. Chem. C* **124**, 26368–26378 (2020).
36. K. Xu, *Chem. Rev.* **114**, 11503–11618 (2014).
37. C. Liu *et al.*, *Sci. Bull.* **65**, 434–442 (2020).
38. K. N. Wood, G. Teeter, *ACS Appl. Energy Mater.* **1**, 4493–4504 (2018).
39. B. A. Rohr, A. R. Singh, J. K. Nørskov, *J. Catal.* **372**, 33–38 (2019).
40. W. F. Castle, *Int. J. Refrig.* **25**, 158–172 (2002).
41. S. Ivanova, R. Lewis, *Chem. Eng. Prog.* **108**, 38–42 (2012).
42. K. Li, S. Z. Andersen, I. Chorkendorff, DTU Data (2021).

ACKNOWLEDGMENTS

We thank the floor managers B. P. Knudsen and P. Strøm-Hansen for helping with the design of the autoclaves and the connection to the mass spectrometer. **Funding:** We gratefully acknowledge the funding by Villum Fonden, part of the Villum Center for the Science of Sustainable Fuels and Chemicals (V-SUSTAIN grant 9455), Innovationsfonden (E-ammonia grant 9067-00010B) and the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement no. 741860). **Author contributions:** Conceptualization: S.Z.A., K.K., I.C., J.K.N. Data curation: K.L., S.Z.A., M.J.S., M.S. Formal Analysis: K.L., S.Z.A., M.J.S., V.J.B., M.S. Investigation: K.L., S.Z.A., M.J.S., V.J.B., M.S. Methodology – Equipment design: S.Z.A., M.S., V.S. Visualization: K.L., S.Z.A., M.J.S., V.J.B. Supervision: I.C., J.K., P.C.K.V., D.C., J.K.N. Writing – original draft: K.L., S.Z.A., M.J.S., V.J.B. Writing – review and editing: K.L., S.Z.A., M.S., V.J.B., K.K., R.S., J.B.P., D.C., J.K., P.C.K.V., J.K.N., I.C. **Competing interests:** A patent application titled "Oxygen enhancement of lithium-mediated electrochemical nitrogen reduction" was submitted on 22 February 2021 regarding the reported increase in FE due to the addition of oxygen mentioned in the paper. Inventors: S.Z.A., K.L., V.J.B., M.S., K.K., R.S., J.B.P., J.K., M.S., P.C.K.V., D.C., J.K.N., and I.C. Institutions: DTU and Stanford University. US provisional application/DK priority-founding application. The authors declare no financial conflicts of interest. **Data and materials availability:** Additional information is provided in the supplementary materials, and the raw data are archived at the Technical University of Denmark (DTU) (*42*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abl4300
Materials and Methods
Supplementary Text
Figs. S1 to S18
Tables S1 and S2
References (*43*–*45*)

9 August 2021; accepted 3 November 2021
10.1126/science.abl4300

REPORTS

PEROVSKITE STRUCTURE

Stabilized tilted-octahedra halide perovskites inhibit local formation of performance-limiting phases

Tiarnan A. S. Doherty[1]†, Satyawan Nagane[1]†, Dominik J. Kubicki[1,2], Young-Kwang Jung[3], Duncan N. Johnstone[4], Affan N. Iqbal[1,5], Dengyang Guo[1,5], Kyle Frohna[1], Mohsen Danaie[6,7], Elizabeth M. Tennyson[1], Stuart Macpherson[1], Anna Abfalterer[1], Miguel Anaya[1,5], Yu-Hsien Chiang[1], Phillip Crout[4], Francesco Simone Ruggeri[8], Sean M. Collins[9], Clare P. Grey[2], Aron Walsh[3,10], Paul A. Midgley[4], Samuel D. Stranks[1,5]*

Efforts to stabilize photoactive formamidinium (FA)–based halide perovskites for perovskite photovoltaics have focused on the growth of cubic formamidinium lead iodide (α-$FAPbI_3$) phases by empirically alloying with cesium, methylammonium (MA) cations, or both. We show that such stabilized FA-rich perovskites are noncubic and exhibit ~2° octahedral tilting at room temperature. This tilting, resolvable only with the use of local nanostructure characterization techniques, imparts phase stability by frustrating transitions from photoactive to hexagonal phases. Although the bulk phase appears stable when examined macroscopically, heterogeneous cation distributions allow microscopically unstable regions to form; we found that these transitioned to hexagonal polytypes, leading to local trap-assisted performance losses and photoinstabilities. Using surface-bound ethylenediaminetetraacetic acid, we engineered an octahedral tilt into pure α-$FAPbI_3$ thin films without any cation alloying. The templated photoactive $FAPbI_3$ film was extremely stable against thermal, environmental, and light stressors.

Although early perovskite solar cells primarily used methylammonium (MA)–based absorber layers, formamidinium (FA)–based perovskites have much greater thermal stability. However, $FAPbI_3$ is challenging both to fabricate and stabilize, because the photoactive cubic phase (α-$FAPbI_3$) consisting of lead iodide octahedra is stable only at temperatures above 150°C, where it is entropically stabilized by the reorienting FA cations (*1*). At room temperature, the energy barrier is readily overcome and the material rapidly transitions to wide-bandgap, face-sharing hexagonal polytypes, such as the 2*H* δ-phase, 4*H*, or 6*H* phases (*2*, *3*). Alloying FA with Cs^+, MA, or both on the A-site cation of the ABX_3 perovskite structure can stabilize photoactive $FAPbI_3$-like cubic structures at room temperature. For example, perovskite solar cells fabricated with $Cs_{0.05}FA_{0.78}MA_{0.17}Pb(I_{0.83}Br_{0.17})_3$ (triple-cation) or those comprising $FAPbI_3$ alloyed with $MAPbBr_3$ perovskites have achieved high power conversion efficiencies (PCEs) with greatly enhanced reproducibility and ambient stability relative to pure $FAPbI_3$ (*4*–*8*). The most successful recent strategies for stabilizing pure $FAPbI_3$ perovskite thin films still incorporate a small fraction of alloying cations, including incorporation of MA through use of methylammonium chloride (in conjunction with formamidinium formate) (*9*), methylammonium thiocyanate vapor (*10*), methylammonium formate (*11*), or other cations such as Cs^+ and methylenediammonium (*12*, *13*).

These approaches lead recent record efficiency tables, and power conversion efficiencies have now exceeded 25.5% in single-junction and 29.5% in tandem configurations (*14*). Nonetheless, degradation to undesirable hexagonal by-products during the lifetime of a photovoltaic panel can still occur (*3*). Nanoscale domains of hexagonal-phase impurities can persist even in high-performing films that appear otherwise cubic in macroscopic measurements (*15*). These trace hexagonal domains induce clusters of deep trap states that are detrimental to performance (*16*, *17*) and induce photodegradation under operational conditions (*15*). Eliminating these hexagonal-phase impurities will be essential for commercial viability of these cells, but doing so requires a fundamental atomic-level understanding of why and how they form.

Improved cubic-phase stability has been attributed to tuning the Goldschmidt tolerance factor toward the perfect cubic perovskite structure through cation mixing (*7*, *18*), templating growth of the corner-sharing cubic structure (*10*, *11*), relaxing strain (*13*), or reducing intrinsic defect density (*9*). We found that stable, photoactive FA-rich perovskites exhibited small, symmetry-breaking, octahedral tilting at room temperature and actually have a noncubic structure. The magnitude of octahedral tilting is very small and only weak superstructure diffraction peaks are created that are below the noise threshold of traditional characterization techniques. We elucidated these features using local, low-dose, nanostructure probes and sensitive photon-counting detectors. This octahedral tilt–stabilized (ots) phase is induced through the alloying of cations and acts as an inherent photoactive material stabilizer by frustrating the transformation from the photoactive noncubic phase to hexagonal wide-bandgap phases. Although this alloying approach provides apparent phase purity if viewed macroscopically, spatial heterogeneity in cation distribution in the film is associated with local nanoscopic regions that are not tilted and thus form residual hexagonal-phase impurities (*15*, *17*).

We demonstrate a strategy in which surface-bound ethylenediaminetetraacetic acid (EDTA) templated the growth of ots-$FAPbI_3$ throughout the bulk film, as elucidated by solid-state nuclear magnetic resonance (NMR) and nuclear quadrupole resonance (NQR) measurements. The ots-$FAPbI_3$ films showed exceptional stability against thermal, atmospheric, and light stressors without any cationic additives.

We first solution-processed thin films of triple-cation $Cs_{0.05}FA_{0.78}MA_{0.17}Pb(I_{0.83}Br_{0.17})_3$ perovskite on SiN transmission electron microscope (TEM) substrates following a previously reported process (*16*). An electron diffraction (ED) pattern (maintaining low electron dose ~10 electrons/$Å^2$) extracted from a scanning ED (SED) measurement (*19*) of the film could in principle be indexed to a $[001]_c$ zone axis of the expected cubic perovskite $Pm\bar{3}m$ structure with a lattice parameter of ~6.3 Å (Fig. 1A) (*7*, *16*). However, in these SED scans with spatial resolution of 5 nm (extracted from individual grains 50 to 200 nm in size), very faint reflections, forbidden from appearing in the $Pm\bar{3}m$ space group, were visible (Fig. 1A, white arrows). The same forbidden reflections were also observed in measurements from many different sample batches and experimental measurements, as well as in analogous pure-iodide $Cs_{0.05}FA_{0.78}MA_{0.17}PbI_3$ thin films (fig. S1). Thus, these triple-cation compositions, regardless of halide composition, have a noncubic structure.

We attributed these superstructure reflections to octahedral tilting whereby the BX_6 corner-sharing octahedra tilt away from perfect cubic symmetry into 1 of 15 lower symmetry

[1]Department of Physics, Cavendish Laboratory, University of Cambridge, Cambridge, UK. [2]Yusuf Hamied Department of Chemistry, University of Cambridge, Cambridge, UK. [3]Department of Materials Science and Engineering, Yonsei University, Seoul, Korea. [4]Department of Materials Science and Metallurgy, University of Cambridge, Cambridge, UK. [5]Department of Chemical Engineering and Biotechnology, University of Cambridge, Cambridge, UK. [6]Electron Physical Science Imaging Centre, Diamond Light Source Ltd., Didcot, UK. [7]Department of Materials, University of Oxford, Oxford, UK. [8]Laboratories of Organic and Physical Chemistry, Wageningen University and Research, Wageningen, Netherlands. [9]School of Chemical and Process Engineering and School of Chemistry, University of Leeds, Leeds, UK. [10]Department of Materials, Imperial College London, London, UK.
*Corresponding author. Email: sds65@cam.ac.uk
†These authors contributed equally to this work.

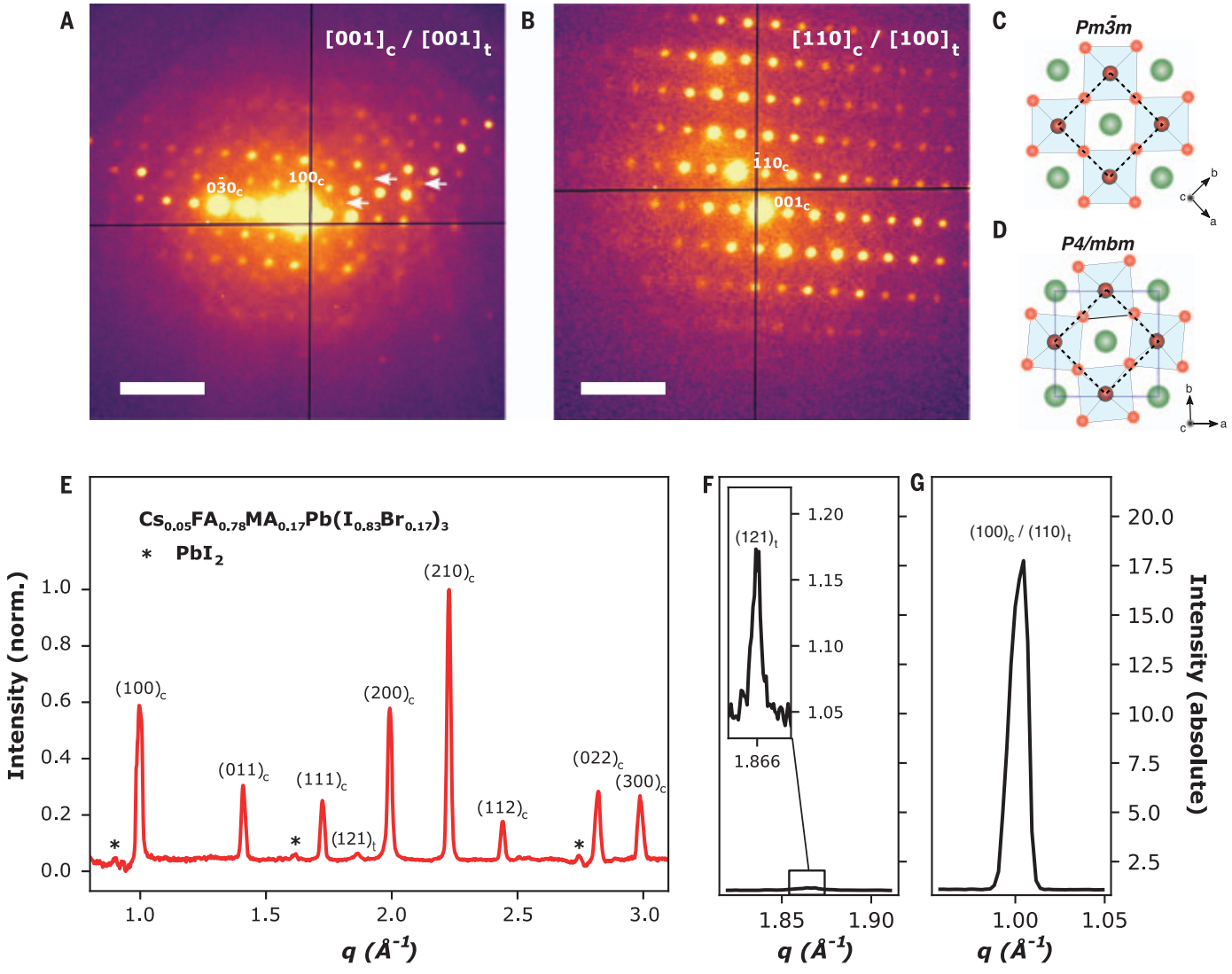


Fig. 1. Structural identification of tetragonal $Cs_{0.05}FA_{0.78}MA_{0.17}Pb(I_{0.83}Br_{0.17})_3$ thin films. (**A**) The ED pattern of a $Cs_{0.05}FA_{0.78}MA_{0.17}Pb(I_{0.83}Br_{0.17})_3$ sample oriented near the [001] zone axis of a cubic unit cell ($[001]_c$) or the $[001]_t$ zone axis of a tetragonal unit cell. The superstructure reflections indicated by white arrows are forbidden from appearing in a cubic structure. (**B**) The ED pattern of a $Cs_{0.05}FA_{0.78}MA_{0.17}Pb(I_{0.83}Br_{0.17})_3$ film oriented near the $[110]_c$/$[100]_t$ zone axis. Superstructure reflections are absent. (**C**) Schematic representation of the cubic *Pm*$\bar{3}$*m* perovskite structure viewed along the $[001]_c$ direction. Green spheres represent A-site cations; red spheres represent halides. B-site cations are represented by gray spheres that are partially visible and are enclosed in blue octahedral cages. The cubic unit cell is indicated by a black dashed box. (**D**) Schematic representation of the tetragonal *P4/mbm* perovskite structure viewed along $[001]_c$. The tetragonal unit cell is indicated by the blue box connecting A-site cations. The pseudo-cubic unit cell is indicated by a black dashed box. (**E**) Mean nano–x-ray diffraction (nXRD) pattern of a $Cs_{0.05}FA_{0.78}MA_{0.17}Pb(I_{0.83}Br_{0.17})_3$ thin film extracted by spatially averaging across a 15 μm × 10 μm region. The pattern was normalized between 0 and 1 by subtracting the minimum value and scaling with the maximum value from the respective map of peak intensity ratio. (**F**) $(121)_t$ tetragonal XRD peak extracted from local nXRD measurements. Inset: Close-up of $(121)_t$ peak. (**G**) $(100)_c$/$(110)_t$ XRD peak extracted from local nXRD measurements. Note that (F) and (G) are plotted on the same intensity scale. Scale bars, 0.5 $Å^{-1}$ [(A) and (B)].

space groups (fig. S2) (*20*). We exclude the possibility that the superstructure reflections could be attributed to effects other than octahedral tilting, such as dynamical scattering, or cation ordering (see text S1). We also exclude the possibility that electron beam-induced structural changes affected the interpretation of our results as our measurements were acquired over an order of magnitude below the electron doses at which beam-induced effects are observed. We also observed these superstructure reflections in halide perovskite compositions known to possess a tilted, noncubic, structure (*21*, *22*) [fig. S3; see also (*19*) and text S1].

Interrogating superstructure reflections at particular orientations allows the unambiguous assignment of space groups from ED patterns of octahedrally tilted perovskites (*20*), and is typically achieved by tilting a single crystalline sample into different orientations and recording an ED pattern at each orientation. However, this methodology is not compatible with the most technologically relevant halide perovskite materials given their polycrystalline nature, small grains, and beam sensitivity. We used the low-dose, large-area scanning capabilities of SED to sample many individual grains at orientations other than $[001]_c$ and determine the actual symmetry of the FA-rich alloyed (triple-cation) perovskite unit cell.

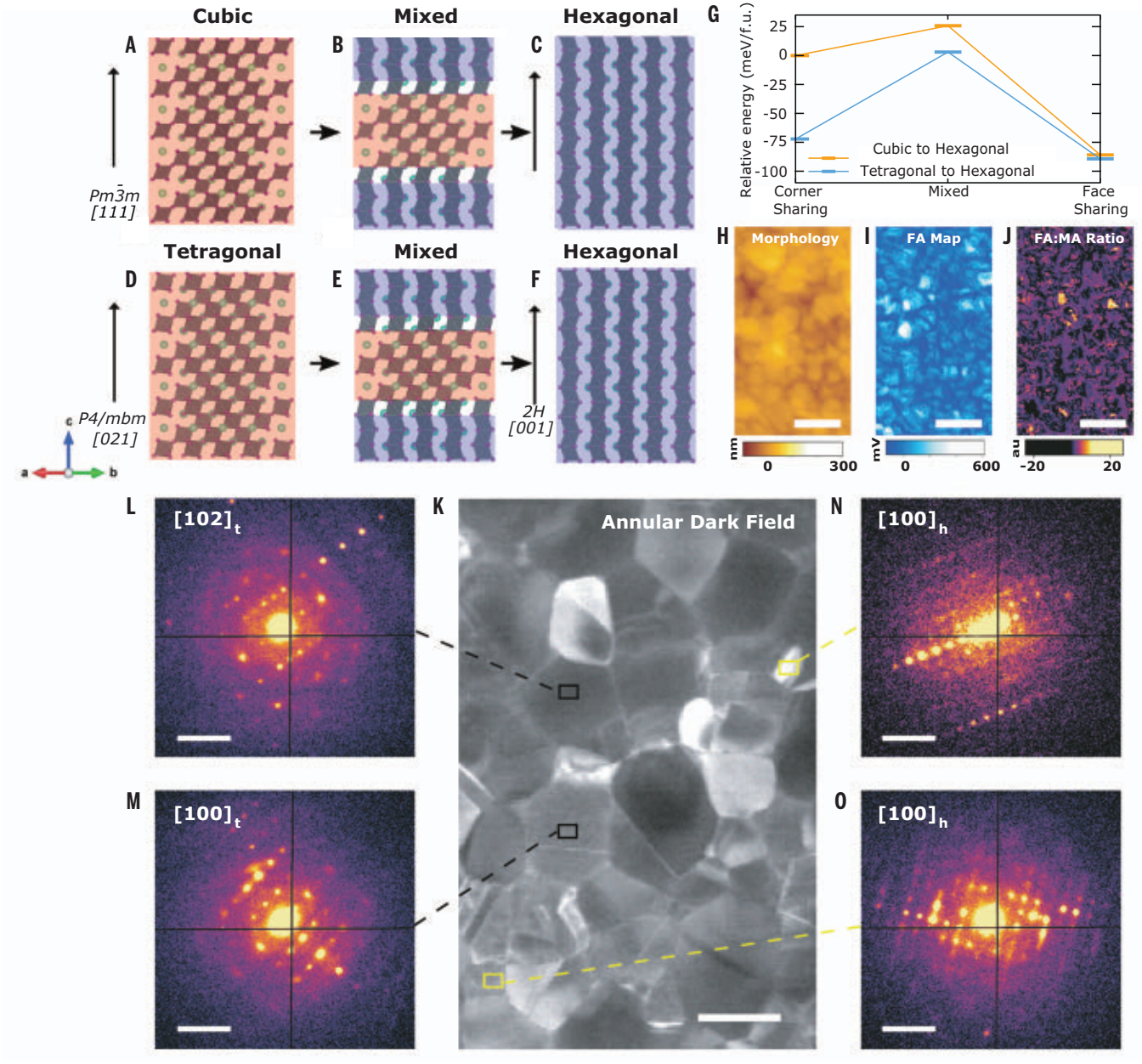


Fig. 2. Tilted octahedra stabilize photoactive perovskites against degradation. (**A** to **C**) Progression of the structural transition from the corner-sharing cubic $Pm\bar{3}m$ structure (A) to the *2H* hexagonal face-sharing structure (C) by way of a mixed corner-sharing and face-sharing intermediary structure (B). (**D** to **F**) Progression of the structural transition from the corner-sharing tetragonal *P4/mbm* structure (D) to the *2H* hexagonal face-sharing structure (F) by way of a mixed corner-sharing and face-sharing intermediate structure (E). (**G**) Relative energy difference between the corner-sharing, mixed, and face-sharing phases for the cubic-to-hexagonal (gold) and tetragonal-to-hexagonal (blue) transition. (**H** and **I**) For a $Cs_{0.05}FA_{0.78}MA_{0.17}Pb(I_{0.83}Br_{0.17})$ thin film, AFM-IR morphology map (H) and IR absorption maps of the FA content (I) (peak at 1712 cm^{-1}). (**J**) IR chemical ratio of the FA:MA cation distribution extracted by dividing the IR map of FA content in (I) by the IR map of MA content in fig. S8. (**K**) Annular dark-field image reconstructed from SED data of a region of a $Cs_{0.05}FA_{0.78}MA_{0.17}Pb(I_{0.83}Br_{0.17})_3$ thin film. Hexagonal phase impurities are shaded in yellow. (**L** and **M**) ED patterns extracted from a black region of interest in (K), revealing a grain oriented near the $[102]_t$ zone axis (L) and a grain oriented near the $[100]_t$ zone axis (M). (**N** and **O**) ED patterns of the yellow boxed region of interest shown in (K), revealing a *2H* hexagonal phase impurity oriented near the $[100]_h$ zone axis (N) and oriented near the $[100]_h$ zone axis (O). Scale bars, 800 nm [(H) to (J)], 200 nm (K), 0.5 $Å^{-1}$ [(L) to (O)].

Group-subgroup relations (fig. S2), along with the symmetry of the experimentally observed superstructure reflections for triple-cation samples, indicated that the most likely space groups were either tetragonal *P4/mbm* or orthorhombic *Pnma* (*20*, *23*). However, the ED pattern of grains oriented near a $\langle 110 \rangle_c$ zone axis (Fig. 1B and figs. S4 to S6) revealed no superstructure reflections when compared to the expected cubic structure. In the *Pnma* space group, superstructure reflections are expected in 2 of 12 $\langle 110 \rangle_c$ zone axis diffraction patterns, whereas in the *P4/mbm* space group, they are expected in none (fig. S4) (*20*). The persistent absence of these reflections in our

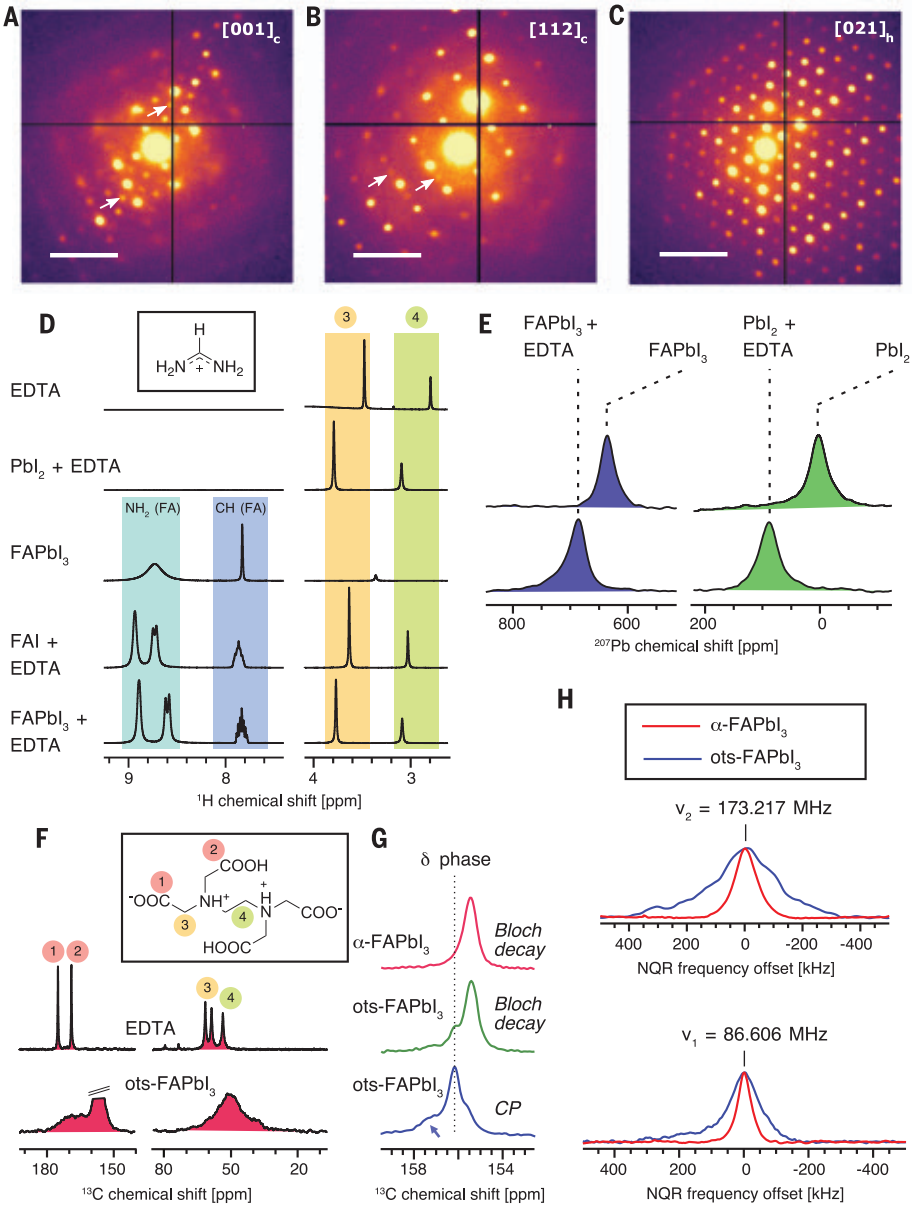


Fig. 3. Surface-bound ethylenediaminetetraacetic acid (EDTA) templates growth of octahedral tilt–stabilized (ots)-$FAPbI_3$. (**A** and **B**) The SED pattern for ots-$FAPbI_3$ perovskite thin films near the $[001]_c$ zone axis (A) and near the $[112]_c$ zone axis (B). A subset of superstructure reflections is highlighted with white arrows in each SED pattern. (**C**) The SED pattern for a hexagonal (6*H*) structure present in the ots-$FAPbI_3$ perovskite thin film oriented near the [021] zone axis. (**D**) Liquid-state ^{1}H NMR of the precursor solution reveals that EDTA interacts with Pb^{2+} ions (in mixtures with PbI_2 and $FAPbI_3$) and FA ions (in mixtures with FAI and $FAPbI_3$). The protonation of FA by EDTA hinders the rotation around the partial double C–N bonds, rendering the two protons within the two NH_2 moieties inequivalent (visible via the peak splitting of the FA NH_2 resonances and the J-coupling on both the FA CH and NH_2 resonances). The inset shows the structure of protonated FA. (**E**) Liquid state ^{207}Pb NMR corroborates the Pb^{2+}-EDTA interaction in the precursor solution of dissolved PbI_2 and $FAPbI_3$. (**F**) ^{1}H-^{13}C CP MAS NMR reveals that EDTA in EDTA-doped $FAPbI_3$ (scraped-off material from drop-cast films) is substantially disordered relative to neat EDTA. The intense peak of FA environments is clipped here for clarity. The inset shows the structure of EDTA. (**G**) The "FA" region of the Bloch decay–detected (i.e., pulse-acquire) ^{13}C MAS NMR spectrum (green) shows that most of the FA environments are unchanged in ots-$FAPbI_3$ and correspond to FA within α-$FAPbI_3$. New environments are also present that correspond to δ-$FAPbI_3$ as well as FA interacting with EDTA (indicated by an arrow in the CP spectrum, blue). (**H**) ^{127}I nuclear quadrupole resonance (NQR) spectra of α-$FAPbI_3$ and ots-$FAPbI_3$ reveal more asymmetric local structure of iodides in the latter material. We attribute this effect to a reduction in overall crystallographic symmetry as a result of structure templating by the surface-bound EDTA. Scale bars, 0.7 $Å^{-1}$ [(A) to (C)].

experimentally observed $\langle 110 \rangle_c$ zone axis patterns thus provides strong evidence that these triple-cation perovskites had a tetragonal *P4/mbm* structure.

The weak nature of the superstructure reflections observed here, which were resolvable with SED given that the interaction of electrons with matter is orders of magnitude stronger than that of x-rays, suggested that the structural deviation from cubic (Fig. 1, C and D) is very small and would be extremely difficult to detect in laboratory Bragg x-ray diffraction (XRD) [see (*19*)], and was consistent with previous assignments of the triple-cation perovskite to a cubic structure (*7*). We used synchrotron-based nano XRD (nXRD) experiments to quantify the degree of octahedral tilting present in FA-rich alloyed samples. A 50-nm focused x-ray probe allowed us to scan local regions of a sample, acquiring both single-crystal patterns from individual grains and powder patterns representative of the bulk of the material by spatially averaging across every probe position in a given scan. Even with monochromated synchrotron light and a highly sensitive photon-counting detector (*19*), we only observed weak diffracted intensity from the $(121)_t$ tetragonal superstructure reflection in a spatially averaged nXRD pattern across a 15 μm × 10 μm region of a triple-cation film sample (Fig. 1E) (*19*). All other tetragonal superstructure peaks had an intensity below or comparable to the noise of the averaged pattern.

However, the local nature of nXRD allowed us to use virtual dark-field imaging (fig. S7) (*19*) to exclusively extract regions in the scan where the Bragg condition was satisfied for tetragonal superstructure peaks. This approach substantially increased the signal-to-noise ratio, allowed extraction of local tetragonal superstructure peaks (Fig. 1F and inset), and mitigated the impact of local film texture on the analysis of diffracted intensity. As the intensity of superstructure peaks relative to primary Bragg peaks is directly related to the degree of octahedral tilting (*24*), we could compare the averaged intensities *I* of the tetragonal $(121)_t$ Bragg peak (Fig. 1F) to the much stronger primary $(100)_c/(110)_t$ Bragg peak (Fig. 1G) to show that $I_{(121)t}$ was 0.75% of $I_{(100)c/(110)t}$. A comparison of this relative intensity to kinematical simulations of a range of different octahedral tilt angles indicated an octahedral rotation in the range 0.75° to 2.75° about the c axis of the

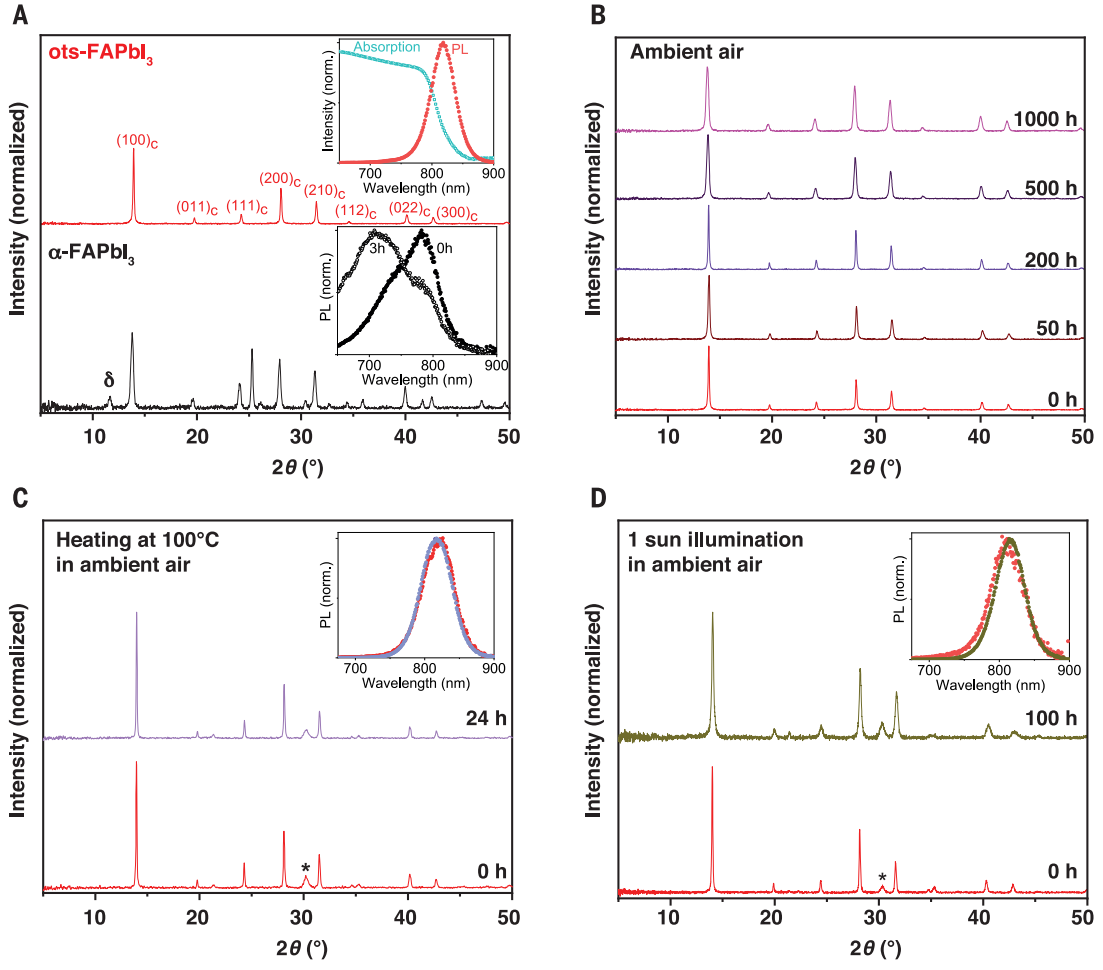


Fig. 4. ots-$FAPbI_3$ perovskite films are highly stable against atmospheric, thermal, and light stressors. (**A**) X-ray diffraction (XRD) pattern for a control α-$FAPbI_3$ sample (bottom pattern, black) taken immediately after exposing to ambient air, with a δ-phase peak already present. The inset shows PL spectra of a film taken initially (0 hours corresponds to ~5 min of total exposure to ambient air) and again after 3 hours of exposure. Note that these control α-$FAPbI_3$ films are rapidly degrading during the measurements in ambient air, and thus these spectra are merely snapshots in time of the samples during the degradation. The top pattern (red) shows the XRD pattern of an ots-$FAPbI_3$ film, indexed to the cubic structure for labeling purposes, with no signature of additional phase impurities. The inset shows the corresponding absorption (open blue symbols) and PL (solid red symbols) spectra of the film. (**B**) XRD pattern of an ots-$FAPbI_3$ film stored in ambient air over a period of 1000 hours, showing negligible change. (**C**) XRD pattern of an ots-$FAPbI_3$ film subjected to continuous heating at 100°C for 24 hours in ambient air, showing very little change. The inset shows that the PL spectra also exhibit minimal change after the heating. (**D**) XRD pattern of an ots-$FAPbI_3$ film subjected to continuous illumination under 1-sun intensity (AM1.5) for 100 hours in ambient air, showing only small changes in patterns. The inset shows that the PL spectra also exhibit minimal change after the illumination, other than a small spectral narrowing. Asterisk in (C) and (D) denotes peak from the ITO substrate.

unit cell (Fig. 1D) [see fig. S7, (*19*), and text S2 for calculation of relative peak intensity, estimation of octahedral rotation and its range, and extraction of local diffraction data]. Our collective results revealed that these FA-rich alloyed cation compositions intrinsically had small octahedral tilting structural distortions that were not observable in macroscopic Bragg diffraction experiments but resolvable by the local Bragg diffraction measurements used here.

The crystallization of FA-rich alloyed perovskites in this slightly tilted, corner-sharing, photoactive phase at room temperature was surprising. We hypothesize that the ots phase led to reported improved stability and resistance to transforming into the hexagonal phases when compared to untreated cubic α-$FAPbI_3$ (*25*). To test whether the tilted octahedra of the tetragonal *P4/mbm* triple-cation perovskite provided an innate barrier to forming a hexagonal face-sharing structure, we probed the thermodynamics of the transformation between corner and face-sharing octahedral networks using first-principles density functional theory (DFT) total energy calculations [see (*19*)]. We considered both the cubic $Pm\bar{3}m$ phase (Fig. 2A) and the tetragonal *P4/mbm* (Fig. 2D) as the starting corner sharing phases and the same hexagonal 2*H* phase as the final face-sharing phase (Fig. 2, C and F). The energy difference (thermodynamic driving force) between the cubic and hexagonal phases was 86 meV (Fig. 2G), whereas the difference between the tetragonal and hexagonal phase was only 17 meV.

We estimated the thermodynamic cost to form a mixed corner/face-sharing phase (Fig. 2B for cubic and Fig. 2E for tetragonal), which is an intermediate between the corner-sharing and face-sharing phases. These values provided a lower bound for the phase transition barrier: We obtain a barrier height of 26 meV per formula unit for the cubic-to-hexagonal phase transition and 75 meV/f.u. for the tetragonal-to-hexagonal phase transition, again indicating that the cubic phase was more susceptible to transitioning to hexagonal polytypes. We note that the phase transition may be further influenced by reorientations of the organic cation (*26*). Although transitions between cubic corner-sharing and hexagonal face-sharing structures are well documented in halide perovskites, oxide perovskites, and silicon carbide materials (*27*–*29*), the same is not true for tetragonal corner-sharing (or other tilted structures) to hexagonal face-sharing transitions.

Given we observed the tetragonal structure in both mixed halide and single halide FA-rich alloyed cation compositions (Fig. 1, A and B, and fig. S1), and that the structurally similar α-$FAPbI_3$ had an average cubic structure at room temperature, as revealed by neutron diffraction experiments (*25*), we propose that room-temperature octahedral tilting in FA-rich perovskites originated primarily from alloying of the FA, Cs^+, or MA cations (or some combination) on the A-site. Specifically, the

mixing of different-sized cations induced a small distortion of the perovskite's local unit cell that frustrated the structural transformation to 2*H* and other hexagonal phases. We note that halide mixing could additionally influence the observed degree of octahedral tilting (fig. S8 and text S1) (*19*).

Using atomic force microscopy–based infrared nanospectroscopy (AFM-IR) to simultaneously map the morphology of the film (Fig. 2H) and the spatial distribution of the chemical signature of the FA (at 1712 cm^{-1}; Fig. 2I) and MA (1466 cm^{-1}; fig. S8) of a triple-cation perovskite thin film, we observed discrete FA-rich regions of the film ~50 to 200 nm in size, as seen in the FA:MA ratio map in Fig. 2J (see also fig. S9) (*30*). SED measurements (Fig. 2K) revealed that although most of the film exhibited *P*4/*mbm* symmetry (Fig. 2, K to M), we also observed discrete inclusions indexable to hexagonal polytype structures ~50 to 150 nm in size and comparable to the length scales of the FA-rich domains (Fig. 2, K, N, and O; see fig. S10 for additional examples).

On the basis of these observations, we concluded that these hexagonal regions were linked to the heterogeneity in the cation distribution. For example, regions of the film with a local excess of FA, such as those visualized by AFM-IR (Fig. 2, I and J), had a reduced content of Cs^+, MA, or both (Fig. 2, J and K, and fig. S9). The locally higher FA content allowed hexagonal phases to form either directly during the crystallization process (which cation alloying would otherwise have inhibited) (*2*, *31*), or indirectly after annealing as the FA-rich environment favored the formation of the photoactive perovskite as a cubic $Pm\bar{3}m$ structure rather than a tetragonal *P4/mbm* structure. The cubic structure then readily transformed to 2*H* and other hexagonal phases.

Even if these phases appear stabilized when probed macroscopically through small additions of other cations (*2*, *32*), typically used film-processing approaches for devices can leave phase impurities that persist on the nanoscale (in general unobservable in macroscopic techniques) (*16*, *17*) and induce degradation under operation (*15*). The pathway to highly stable and efficient FA-rich perovskite devices is through the slight distortion of octahedra across the sample to a degree that frustrates the transition from corner-sharing to face-sharing structures but does not compromise optoelectronic properties for example by widening the bandgap or reducing carrier lifetimes (*33*). To selectively induce octahedral tilt in $FAPbI_3$ at room temperature without alloying multiple A-site cations, we used ethylenediaminetetraacetic acid (EDTA) as an additive to the precursor solution of $FAPbI_3$ (*19*), considering the potential for such bifunctional molecules to interact with both Pb^{2+} ions and ammonium cations in precursor solutions, during film formation, or both. We spin-coated EDTA-containing perovskite precursors on substrates that were then annealed at 150°C for 1 hour in a nitrogen glove box to form a visibly stable, optically active film of the black perovskite phase (see fig. S11 for optimization of EDTA concentration). We did not include any other additives, which allowed us to study the pure $FAPbI_3$ systems.

SED patterns extracted from the resulting perovskite thin films (Fig. 3, A and B) confirmed the presence of superstructure reflections identical to those observed in the *P*4/*mbm* FA-rich perovskites, indicating that octahedral tilting occurred in the fabricated $FAPbI_3$. Occasionally, superstructure reflections inconsistent with *P*4/*mbm* symmetry were observed (fig. S12), which suggested that multiple octahedral tilt systems were present that we were unable to unambiguously assign. Notably, in each place that we observed a corner-sharing photoactive perovskite structure, we also observed octahedral tilting in ots-$FAPbI_3$. Furthermore, in the infrequent local regions where octahedral tilting was not observed, hexagonal-phase impurities were present (Fig. 3C). These results provided further support for our assertion that octahedral tilt is critical for minimizing the formation of such phase impurities.

To elucidate the microscopic mechanism of the EDTA-induced stabilization of $FAPbI_3$, we performed NMR experiments to probe local structure (*34*, *35*). ^1H liquid-state NMR of the precursor solutions showed a prominent shift of the acetate and ethylenic CH_2 groups of EDTA, added in the neat acid form (*36*), in the presence of dissolved PbI_2 at $\Delta\delta$ = +0.3 parts per million (ppm) and $FAPbI_3$ at $\Delta\delta$ = +0.3 ppm (Fig. 3D). Although EDTA is a strong chelator, we did not observe the formation of a long-lived, hexadentate Pb-EDTA chelate that would lead to splitting of the acetate methylene protons (see text S3) (*37*). We attributed the relative shifts to a combination of changes in the protonation equilibrium because EDTA can exist in six forms with different protonation levels, and a shorter-lived or more disordered Pb^{2+}-EDTA complex. Additional proof of the Pb^{2+}-EDTA interaction was provided by the relative shift in ^{207}Pb NMR spectra of the precursor solutions when EDTA is added (Fig. 3E). In addition, EDTA protonates the FA moiety, hindering the C–N bond rotation, as indicated by the appearance of a set of signals associated with the two inequivalent NH_2 groups of FA and the corresponding CH multiplet (*38*).

We next investigated ots-$FAPbI_3$ by solid-state NMR. ^{207}Pb and ^{14}N solid-state NMR spectra of scraped-off material from drop-cast films showed that the perovskite component of ots-$FAPbI_3$ was virtually identical to that of control α-$FAPbI_3$, within the sensitivity of these two techniques (fig. S13). ^{13}C solid-state NMR allowed us to elucidate the speciation of EDTA in the ots-$FAPbI_3$ material (Fig. 3F). Although neat crystalline EDTA is characterized by narrow ^{13}C resonances [full width at half maximum (FWHM) = 0.3 to 0.9 ppm], the FWHM of both the carbonyl and methylene carbons in ots-$FAPbI_3$ was 14 to 15 ppm, indicating that the EDTA was in an amorphous phase (or component). The ^{13}C resonances of FA revealed that there were multiple FA-containing local environments. Although the largest component corresponded to the three-dimensional perovskite phase of ots-$FAPbI_3$, there was also a smaller component corresponding to the residual δ-phase, consistent with the presence of small fractions of hexagonal polytypes in the SED data (Fig. 3C).

We also detected another substantially broader FA peak corresponding to a more disordered FA local environment, which we attribute to the interfacial FA ions of the perovskite phase interacting directly with EDTA (Fig. 3G, arrow). Cross-polarization (CP) indicates that the local environment was rigid and not undergoing rapid near-isotropic reorientation characteristic for FA inside the A-site cation cage. ^{14}N NMR is highly sensitive to changes in lattice symmetry induced by incorporation of additives into the perovskite structure (*35*). The ^{14}N spectrum of the fast-reorienting FA inside a 3D perovskite cage (fig. S13) showed that the symmetries of ots-$FAPbI_3$ and control α-$FAPbI_3$ are essentially identical within the sensitivity of this approach, indicating that the EDTA did not incorporate into the perovskite structure.

To elucidate the much smaller effect of EDTA on the lattice symmetry, we used ^{127}I NQR. In NQR, the nuclear energy levels are split by the electric field gradient (EFG) around the nucleus and not by an external magnetic field, as in NMR (fig. S13). The resulting transitions can be driven at specific frequencies that depend on the magnitude of the EFG, which is determined by local symmetry. The NQR spectrum of the control α-$FAPbI_3$ sample contained two signals corresponding to the $\pm3/2 \longleftrightarrow \pm5/2$ and $\pm1/2 \longleftrightarrow \pm3/2$ transitions of the ^{127}I nucleus at 173.217 and 86.606 MHz, with FWHM of 87 and 47 kHz, respectively (Fig. 3H). The broadening of the ^{127}I NQR resonances in ots-$FAPbI_3$ (FWHM = 246 and 112 kHz for $\pm3/2 \longleftrightarrow \pm5/2$ and $\pm1/2 \longleftrightarrow \pm3/2$, respectively) evidenced that there was a broader distribution of local iodide environments compared to the control α-$FAPbI_3$. This result was consistent with the resulting octahedrally tilted phase having lower symmetry.

In a tetragonal $FAPbI_3$, each of the two NQR resonances present in cubic $FAPbI_3$ would split into two because there are two crystallographically inequivalent iodide sites in the unit cell of tetragonal $FAPbI_3$ (*39*). Splitting was not observed here, indicating that the

degree of distortion is characterized by a distribution rather than a clear-cut value throughout the material, consistent with SED observations of multiple local octahedral tilting symmetries (fig. S12). The FWHM of the $\pm 3/2 \leftrightarrow \pm 5/2$ transition of ots-$FAPbI_3$ (246 kHz) corresponds to the maximum degree of distortion present in neat tetragonal β-$FAPbI_3$ at ~280 K (*40*). Notably, ^{14}N NMR (fig. S13), which has recently been used to show that the symmetry of A-site cation cages of $FAPbI_3$ increases when it is stabilized with methylammonium thiocyanate (*10*), was not sensitive enough to detect the minor deviations from cubic symmetry identified here by NQR, SED, and nXRD (text S4).

Taken together, these results show that we induced octahedral distortion in the material through a structure-directing effect of EDTA, with the EDTA binding to the $FAPbI_3$ surface but not incorporating into the $FAPbI_3$ structure. We expect that this will spur further experimental and computational work to establish the exact binding modes of EDTA, identify other growth-templating additives, and understand their effect on crystallization and the resulting spatial variation of the tilt.

We examined octahedral tilt stabilization of the $FAPbI_3$ films by subjecting them to a variety of external stressors. The bulk XRD pattern of the ots-$FAPbI_3$ film (Fig. 4A, red) could in principle be indexed to a $Pm\bar{3}m$ cubic structure, despite possessing a lower-symmetry tilted structure (see Fig. 3, A and B), as the superstructure reflections arising from octahedral tilting were below the detection limit of bulk XRD measurements. There was no macroscopic evidence for phase impurities, despite trace amounts on the nanoscale (see Fig. 3C), consistent with the clean absorption and photoluminescence (PL) spectra observed (Fig. 4A, top inset). However, the control α-$FAPbI_3$ films showed the macroscopic presence of the 2*H* δ-phase after only ~5 min (time required to load and measure the sample) of air exposure (XRD peak at 11.6°; Fig. 4A, black), together with an asymmetric PL peak (Fig. 4A, inset; see fig. S14 for cleaner PL spectra of encapsulated control samples). Furthermore, the PL lifetime of the control α-$FAPbI_3$ film was reduced by a factor of 4 with respect to an ots-$FAPbI_3$ film (fig. S14), consistent with the phase impurities in the control films acting as nonradiative recombination centers (*17*). Exposure of the control film to ambient air for 3 hours caused the PL spectral shape and position to change rapidly and substantially (Fig. 4A, bottom inset). These changes were concomitant with further growth of the intensity of the 2*H* δ-phase peak (fig. S15).

After 1000 hours in ambient air, XRD patterns of the ots-$FAPbI_3$ films did not show any macroscopic conversion to hexagonal phases (Fig. 4B), and we observed the presence of impurity phases only after 1500 hours of air exposure (fig. S16). Furthermore, we observed similar phase stability when heating the ots-$FAPbI_3$ perovskite films in ambient air for 24 hours at 100°C, with negligible change in the XRD pattern (Fig. 4C; similar results in nitrogen, fig. S17) or PL spectra (Fig. 4C, inset) after this extended heating. After continuous illumination of the ots-$FAPbI_3$ film under 1-sun (AM1.5) intensity for 100 hours in ambient air, there were no phase impurities evident in XRD (Fig. 4D) or sizable changes observed in PL spectra (Fig. 4D, inset) other than a slight narrowing of the peak and a very slight redshift, the origin of which is currently unclear. The PL properties of the ots-$FAPbI_3$ films were actually enhanced after the heating and illumination tests: The PL lifetimes reached 83 ns after the heating (fig. S18) and 592 ns after the illumination (fig. S19), consistent with light- and oxygen-assisted passivation reported previously for halide perovskites (*40*). These combined results show the resilient stabilizing effect against the generation of impurity phases that the octahedral tilting imparts to photoactive FA-rich perovskites, even for bare films in ambient air under rigorous external stressors.

We have shown that the intrinsic stabilization mechanism of FA-rich mixed-cation systems is an octahedral tilt induced by cation alloying. This octahedral tilting is so minor (~2°) as to be undetectable with bulk characterization techniques yet, remarkably, frustrates the transformation from photoactive, tilted perovskite phases to wide-bandgap, performance-limiting hexagonal polytypes (e.g., the 2*H* δ-phase). We propose that the recent reports of stabilized cubic α-$FAPbI_3$, which have produced devices that are leading the efficiency tables and make use of constituent cations and other additives to improve stability, serendipitously benefit from this same minor octahedral tilting (*9*–*13*). However, homogeneously inducing a tilted structure through cation alloying across a perovskite film is already challenging at lab scale and will only become more so at commercial scale. Any local regions (even trace amounts) of a fabricated FA-rich perovskite that do not possess a tilted structure and are thus cubic will more readily transform to hexagonal phases, which generates nonabsorbing material as well as deep traps and photodegradation pathways under operation (*15*, *17*).

Developing new strategies that can work both in conjunction with cation-alloying approaches and independently of them to homogenize nanoscale phase stability and eliminate residual traps will be critical to realize single-junction and tandem perovskite photovoltaics operating near their performance limits throughout their commercial life cycle (*41*). This is especially true for the most promising compositions for single-junction commercialization, such as α-$FAPbI_3$, where cationic additives produce unwanted shifts to higher bandgap and compromise thermal stability (*6*, *7*, *9*). Here, we have outlined key guidelines for achieving this, by templating the growth of octahedral tilting through additives that do not incorporate into the perovskite structure, such as EDTA, without the use of additional A-site cations.

REFERENCES AND NOTES

1. T. Chen *et al.*, *Sci. Adv.* **2**, e1601650 (2016).
2. P. Gratia *et al.*, *ACS Energy Lett.* **2**, 2686–2693 (2017).
3. P. E. Marchezi *et al.*, *J. Mater. Chem. A* **8**, 9302–9312 (2020).
4. W. S. Yang *et al.*, *Science* **348**, 1234–1237 (2015).
5. W. S. Yang *et al.*, *Science* **356**, 1376–1379 (2017).
6. N. J. Jeon *et al.*, *Nature* **517**, 476–480 (2015).
7. M. Saliba *et al.*, *Energy Environ. Sci.* **9**, 1989–1997 (2016).
8. T. Bu *et al.*, *Science* **372**, 1327–1332 (2021).
9. J. Jeong *et al.*, *Nature* **592**, 381–385 (2021).
10. H. Lu *et al.*, *Science* **370**, eabb8985 (2020).
11. W. Hui *et al.*, *Science* **371**, 1359–1364 (2021).
12. H. Min *et al.*, *Science* **366**, 749–753 (2019).
13. G. Kim *et al.*, *Science* **370**, 108–112 (2020).
14. NREL, Best Research-Cell Efficiency Chart (2021); www.nrel.gov/pv/cell-efficiency.html.
15. S. Macpherson *et al.*, Local Nanoscale Defective Phase Impurities Are the Sites of Degradation in Halide Perovskite Devices. arXiv 2107.09549 [cond-mat] (20 July 2021).
16. T. A. S. Doherty *et al.*, *Nature* **580**, 360–366 (2020).
17. S. Kosar *et al.*, *Energy Environ. Sci.* 10.1039/D1EE02055B (2021).
18. J. Y. Kim, J.-W. Lee, H. S. Jung, H. Shin, N.-G. Park, *Chem. Rev.* **120**, 7867–7918 (2020).
19. See supplementary materials.
20. D. I. Woodward, I. M. Reaney, *Acta Crystallogr. B* **61**, 387–399 (2005).
21. R. E. Beal *et al.*, *Matter* **2**, 207–219 (2020).
22. S. Chen *et al.*, *Nat. Commun.* **9**, 4807 (2018).
23. R. dos Reis *et al.*, *Appl. Phys. Lett.* **112**, 071901 (2018).
24. G. O. Jones, P. A. Thomas, *Acta Crystallogr. B* **58**, 168–178 (2002).
25. M. T. Weller, O. J. Weber, J. M. Frost, A. Walsh, *J. Phys. Chem. Lett.* **6**, 3209–3212 (2015).
26. E. M. Mozur *et al.*, *Chem. Mater.* **29**, 10168–10177 (2017).
27. L. S. Ramsdell, *Am. Mineral.* **32**, 64–82 (1947).
28. C. C. Stoumpos, C. D. Malliakas, M. G. Kanatzidis, *Inorg. Chem.* **52**, 9019–9038 (2013).
29. S. Fop *et al.*, *Nat. Mater.* **19**, 752–757 (2020).
30. R. Szostak *et al.*, *Sci. Adv.* **5**, eaaw6619 (2019).
31. H. T. Pham *et al.*, *Nano Energy* **87**, 106226 (2021).
32. H. X. Dang *et al.*, *Joule* **3**, 1746–1764 (2019).
33. R. Prasanna *et al.*, *J. Am. Chem. Soc.* **139**, 11117–11124 (2017).
34. L. Piveteau, V. Morad, M. V. Kovalenko, *J. Am. Chem. Soc.* **142**, 19413–19437 (2020).
35. D. J. Kubicki, S. D. Stranks, C. P. Grey, L. Emsley, *Nat. Rev. Chem.* **5**, 624–645 (2021).
36. S. Aime, R. Gobetto, R. Nano, E. Santucci, *Inorg. Chim. Acta* **129**, L23–L25 (1987).
37. E. Hafer *et al.*, *Magn. Reson. Chem.* **58**, 653–665 (2020).
38. W. T. M. Van Gompel *et al.*, *J. Phys. Chem. C* **122**, 4117–4124 (2018).
39. K. Yamada *et al.*, *Bull. Chem. Soc. Jpn.* **91**, 1196–1204 (2018).
40. R. Brenes *et al.*, *Joule* **1**, 155–167 (2017).
41. Y. Deng *et al.*, *Nat. Energy* **6**, 633–641 (2021).

ACKNOWLEDGMENTS

We acknowledge J. Parker and P. Quinn for support during experiments on the I14 beamline at Diamond Light Source. We thank Diamond Light Source for access and support in use of beamline I14 (proposal number sp20420) and the electron Physical Science Imaging Centre (ePSIC; Instrument E02 and proposal number MG24111) that contributed to the results presented here. Via our membership of the UK's HEC Materials Chemistry Consortium, which is funded by the Engineering and Physical Sciences Research Council (EPSRC; EP/L000202), this work used the ARCHER UK National Supercomputing Service (www.archer.ac.uk). **Funding:** Supported by a National University of Ireland Travelling Studentship (T.A.S.D.); a Newton International

Fellowship funded by the Royal Society and SERB (S.N.); the Royal Society and Tata Group (UF150033) (S.D.S.); EPSRC grant EP/R008779/1 (P.A.M.); the European Research Council under the European Union's Horizon 2020 research and innovation program (HYPERION grant agreement 756962); EPSRC grant EP/R023980/1; a George and Lilian Schiff Studentship, Winton Studentship, EPSRC studentship, Cambridge Trust Scholarship, and Robert Gardiner Scholarship (K.F.); Marie Skłodowska-Curie actions (grant agreements 841136, 841265, and 841386, respectively) under the European Union's Horizon 2020 research and innovation program (D.J.K., E.M.T., and M.A.); scholarships from the British Spanish Society and the Sir Richard Stapley Educational Trust (A.N.I.); an EPSRC studentship (S.M.); the Royal Society (A.A.); a University of Leeds academic fellowship (S.M.C.); an EPSRC studentship (P.C.); EU ERC Advanced Fellowship DLV-835073 (C.P.G.); National Research Foundation of Korea grant 2018R1C1B6008728; and the European Union's Horizon 2020 INFRAIA program (ESTEEM3–grant agreement 823717). **Author contributions:** Conceptualization: T.A.S.D., S.N., S.D.S. Methodology: T.A.S.D., S.N., D.J.K., D.N.J., C.P.G., A.W., P.A.M., S.D.S. Investigation: T.A.S.D., S.N., D.J.K., Y.-K.J., D.N.J., K.F., F.S.R., D.G., A.N.I., E.M.T., Y.-H.C., P.C., M.A., S.M., M.D., S.M.C., A.A. Visualization: T.A.S.D., S.N., D.J.K., S.D.S. Funding acquisition: S.D.S., A.W., C.P.G., P.A.M. Project administration: SDS. Supervision: S.D.S., A.W., C.P.G., P.A.M. Writing–original draft: T.A.S.D., S.N., D.J.K., S.D.S. Writing–review and editing: T.A.S.D., S.N., D.J.K., Y.-K.J., D.N.J., K.F., F.S.R., D.G., A.N.I., E.M.T., Y.-H.C., P.C., M.A., S.M., A.A., M.D., S.M.C., C.P.G., A.W., P.M., S.D.S. **Competing interests:** S.D.S. is a co-founder of Swift Solar Inc. **Data and materials availability:** The data that support the findings of this study are available from the corresponding author upon request and at the Apollo repository (DOI: 10.17863/CAM.78426). All (other) data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abl4890
Materials and Methods
Supplementary Text
Figs. S1 to S19
Table S1
References (*42*–*64*)

27 July 2021; accepted 5 November 2021
10.1126/science.abl4890

ZEOLITES

A stable aluminosilicate zeolite with intersecting three-dimensional extra-large pores

Qing-Fang Lin[1]†, Zihao Rei Gao[2,3]†, Cong Lin[4,5]†, Siyao Zhang[1], Junfeng Chen[6], Zhiqiang Li[6], Xiaolong Liu[7], Wei Fan[8], Jian Li[2,4,9]*‡, Xiaobo Chen[6]*, Miguel A. Camblor[3]*, Fei-Jian Chen[1]*

Zeolites are crystalline porous materials with important industrial applications, including uses in catalytic and adsorption-separation processes. Access into and out of their inner confined space, where adsorption and reactions occur, is limited by their pore apertures. Stable multidimensional zeolites with larger pores able to process larger molecules are in demand in the fine chemical industry and for the oil processing on which the world still relies for fuels. Currently known extra-large-pore zeolites display poor stability and/or lack pore multidimensionality, limiting their usefulness. We report ZEO-1, a robust, fully connected aluminosilicate zeolite with mutually intersecting three-dimensional extra-large plus three-dimensional large pores. ZEO-1 is stable up to 1000°C, has an extraordinary specific surface area (1000 square meters per gram), and shows potential as a catalytic cracking catalyst.

Zeolites with multidimensional interconnected extra-large pores—pores limited by rings of more than 12 SiO_4 or AlO_4 members [>12-membered rings (12MRs)]—are in demand to allow the processing of large molecules (*1*). In the past three decades, 28 types of zeolites with extra-large pores have been successfully synthesized (*2*), but several factors limit their applicability. First, most of these zeolites require germanium (Ge) to be synthesized, which increases their cost and severely decreases their thermal and hydrothermal stability (*3*, *4*). Second, many are "interrupted frameworks" rather than true zeolites. Such frameworks contain some tetrahedra that do not share all of their vertices, which can decrease robustness (*5*, *6*), and contain dangling OH groups that typically reduce the pore space (*7*). The only reported zeolites with multidimensional extra-large pores (-**ITV**, -**CLO**, -**IFU**, -**IRY**, and -**IFT**) are all interrupted germanosilicates or gallium phosphates with intrinsically low stability (*8*). Among the stable fully connected aluminosilicate zeolites, none possess a multidimensional system of extra-large pores but at most large 12MR pores in topologies discovered decades ago (***BEA**, **FAU**, and **EMT**) (*2*). The recently reported aluminosilicates PST-32 and PST-2 with known **SBS** and **SBS**/**SBT** topologies have an aluminosilicate composition that ensures a good thermal stability, and their pore architecture similar to **FAU** allows them to compete with this successful catalyst (*9*). The long lack of further improvements in the development of more open stable zeolites might suggest that there could be some sort of fundamental limit to the porosity exhibited by aluminosilicates.

We report ZEO-1 (zeolite number 1 of Anhui ZEO New Material Technology Co.), an aluminosilicate zeolite with a multidimensional system of interconnected extra-large pores. This aluminosilicate has a high silica content (Si/Al = 14.5) and a zeolitic, noninterrupted framework, with high thermal and hydrothermal stability. The pore system of ZEO-1 contains both extra-large three-dimensional (3D) 16MR and large 3D 12MR channels with high interconnectivity that results in three types of supercages with four windows of 16MR and/or 12MR. Two of these supercages are larger than those found in **FAU**, **EMT**, and PST-32 and PST-2 (table S1). These features make ZEO-1 one of the stable zeolites with lowest framework density and ultrahigh surface area. Its open framework with large cavities accessible through 16MR windows along the three perpendicular directions (constituting a 3D extra-large 16MR pore system) and the presence in its pores of active acid sites enable good heavy-oil conversion and selectivity to fuels and commodities [gasoline, diesel, and liquified petroleum gas (LPG)] during fluid catalytic cracking (FCC), competing with the performance of the highly optimized 12MR zeolite USY (ultra-stable Y zeolite; **FAU** topology), which is the current workhorse of this reaction.

We used tricyclohexylmethylphosphonium (TCyMP) as the organic structure-directing agent (OSDA) for ZEO-1 (fig. S1), whose complex structure was determined from nanosized crystals (<200 nm) by means of electron diffraction (3D ED) and refined against synchrotron powder x-ray diffraction (SPXRD), revealing that ZEO-1 crystallized in a body-centered tetragonal unit cell [with edge dimensions (a) = 43.3056(4) Å and (c) = 25.0010(8) Å; numbers in parentheses indicate estimated standard deviations] with Laue symmetry 4/*mmm*. The structure determination was a great challenge because of the small crystal size (50 to 200 nm) (fig. S2), large unit cell, and weakly scattering elements. We applied low-dose illumination and fast data acquisition to collect 3D ED data with continuous rotation electron diffraction (cRED) to

[1]Department of Chemistry, Bengbu Medical College, Bengbu 233030, China. [2]Anhui ZEO New Material Technology Co., 778 Dongliu Road, Hefei 230071, China. [3]Instituto de Ciencia de Materiales de Madrid, Consejo Superior de Investigaciones Científicas (ICMM-CSIC), c/Sor Juana Inés de la Cruz 3, 28049 Madrid, Spain. [4]College of Chemistry and Molecular Engineering, Peking University, Beijing, China. [5]Department of Mechanical Engineering, The Hong Kong Polytechnic University, Kowloon, Hong Kong, China. [6]State Key Laboratory of Heavy Oil Processing, China University of Petroleum, Qingdao 266580, China. [7]State Key Laboratory of Optoelectronic Materials and Technologies, School of Materials, Sun Yat-Sen University, Guangzhou 510275, China. [8]Department of Chemical Engineering, University of Massachusetts, Amherst, MA 01003, USA. [9]Berzelii Center EXSELENT on Porous Materials, Department of Materials and Environmental Chemistry, Stockholm University, 10691 Stockholm, Sweden.
***Corresponding author. Email: cfj-cumt@163.com (F.-J.C.); macamblor@icmm.csic.es (M.A.C.); chenxiaobo@upc.edu.cn (X.C.); jxpxlijian@pku.edu.cn (J.L.)**
†These authors contributed equally to this work.
‡Present address: Department of Fibre and Polymer Technology School of Engineering Sciences in Chemistry, Biotechnology and Health KTH Royal Institute of Technology, Teknikringen 56-58, SE-100 44 Stockholm, Sweden.

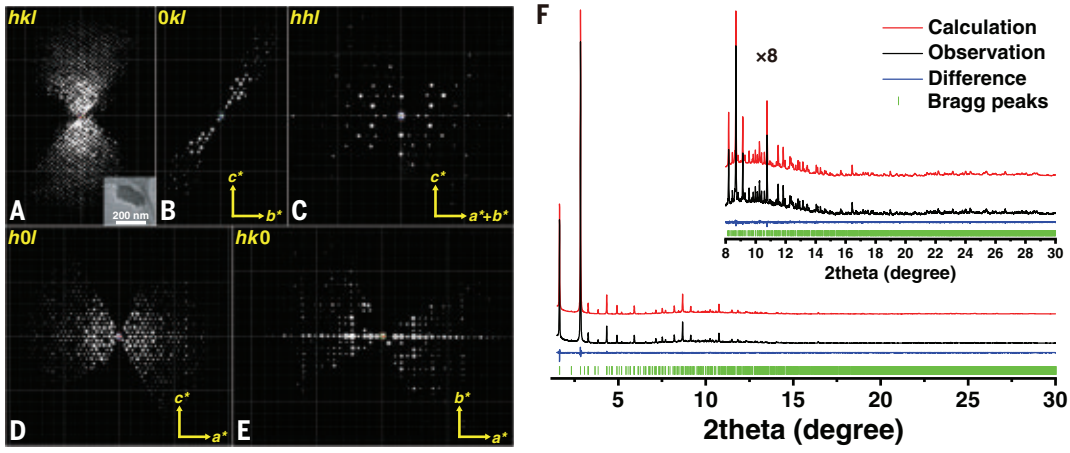


Fig. 1. Structure solution and Rietveld refinement of ZEO-1. (**A**) Projection of 3D reciprocal lattice reconstructed from a typical cRED dataset. (Inset) The crystal (ZEO-1) from which data were collected. (**B** to **E**) The 2D slices of reflections (B) 0*kl*, (C) *hhl*, (D) *h*0*l*, and (E) *hk*0 cut from the reconstructed reciprocal lattice. The reflection conditions are *hkl*, $h + k + l = 2n$; 0*kl*, $k + l = 2n$; *hhl*, $2h + l = 4n$; *hk*0, $h = 2n$, $k = 2n$; and *h*0*l*, $l = 4n$, where n is a natural number. (**F**) Rietveld refinement plots of calcined ZEO-1, with profile, weighted profile, and expected profile fitting factors $R_p = 0.0279$, $R_{wp} = 0.0337$, and $R_{exp} = 0.0241$, respectively.

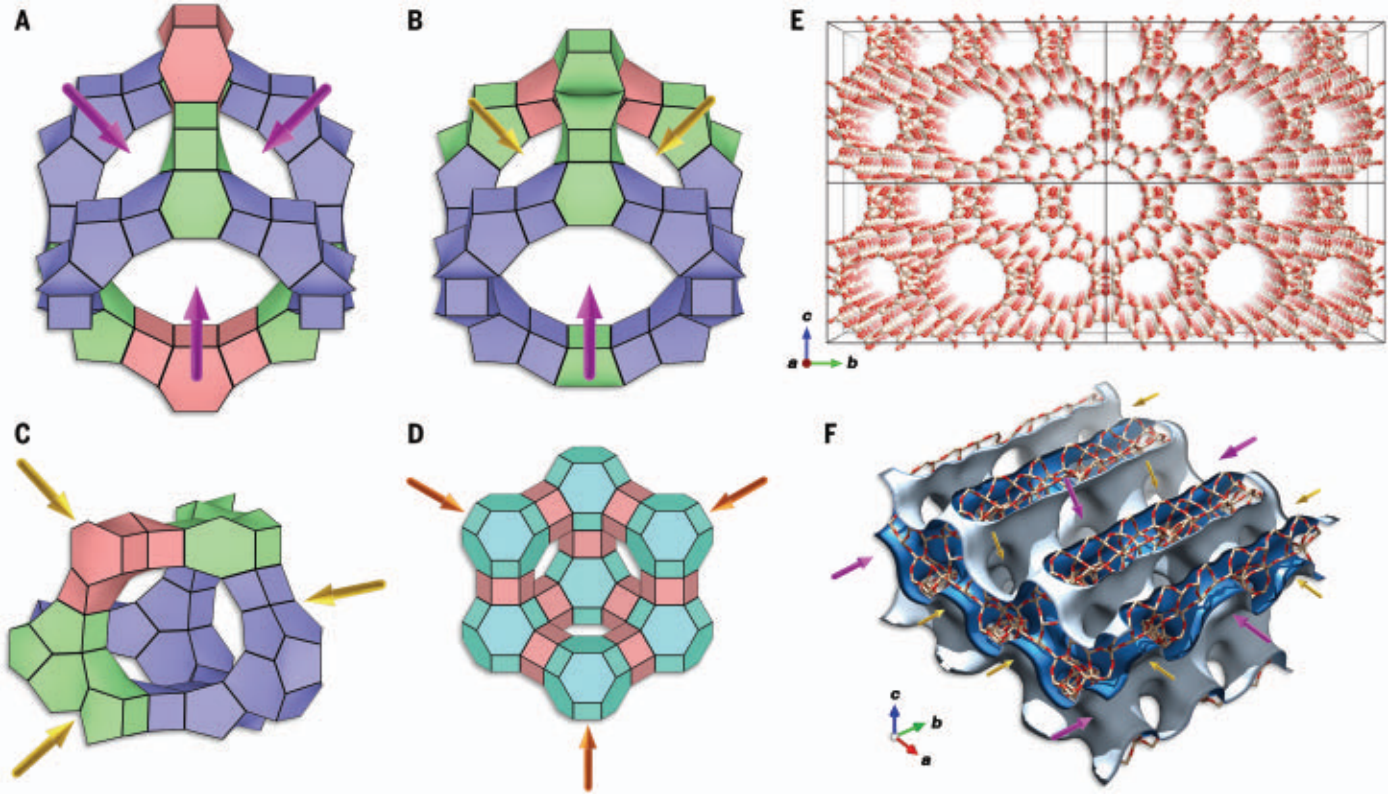


Fig. 2. The three supercages and channel system in ZEO-1. (**A** to **D**) The pore system contains three supercages with (A) 16MR, (B) 16-12MR, and (C) 12MR apertures that are compared here with (D) **FAU** 12MR supercage. (**E**) Structure viewed along *a* axis and (**F**) the 3D channel system with the structure model superimposed. Purple arrows, 16MR; yellow arrows, 12MR; orange ball and stick, silicon; red ball and stick, oxygen.

overcome the difficulties of small crystals and electron-beam damage on the structure determination. The systematic extinctions suggested that the space group was $I4_1/amd$ (#141). The reconstructed reciprocal lattice from a typical dataset is shown in Fig. 1, A to E.

The structure of ZEO-1 calcined at 600°C (3 hours ramp, 6 hours plateau) was successfully solved from cRED data [95% completeness with a resolution of 0.81 Å and a merging residual factor (R_{int}) value of 0.3559] (table S3) through an ab initio method with SHELXLT software (*10*). The framework was refined anisotropically against the cRED data without any restraints, and the refinement converged to an unweighted residual factor (R_1) value of 0.1468 and goodness of fit (GOF) of 0.947 (table S3). We performed further Rietveld refinement of the calcined structure against SPXRD data (Fig. 1F and table S7), which resulted in more precise bond lengths and angles (tables S8 and S9). Additionally, Rietveld refinement of the as-made material (fig. S15 and tables S4 to S6) revealed that the disordered

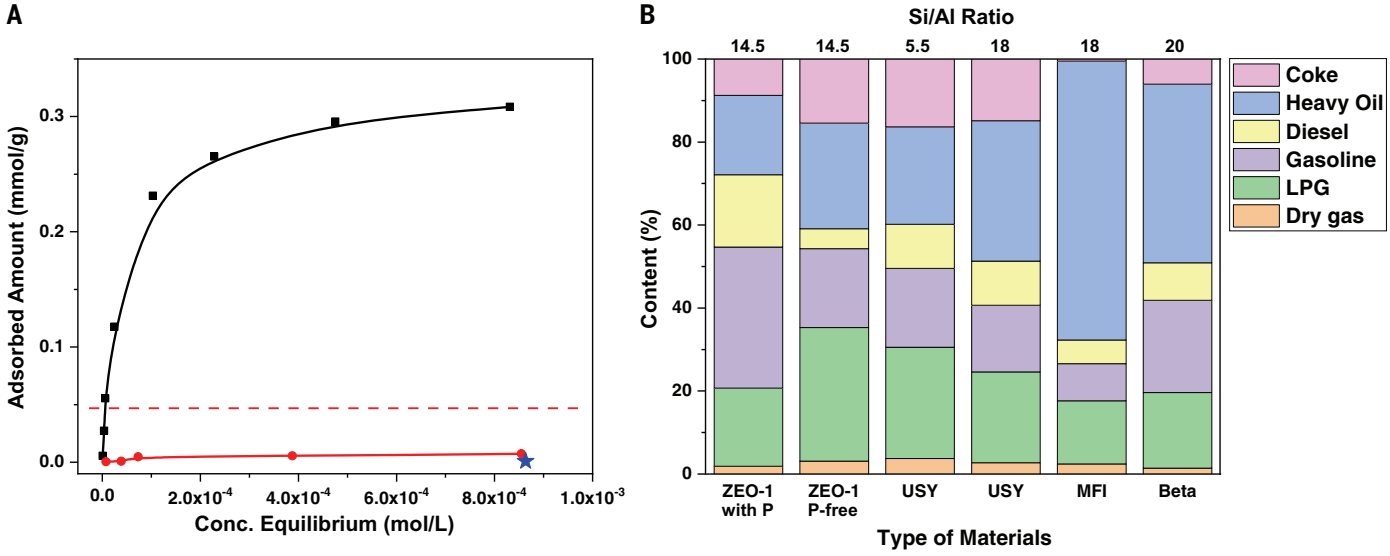


Fig. 3. The extra-large pores of ZEO-1 allow entrance and processing of large molecules. (**A**) Nile red adsorption isotherms (equilibrium concentration versus adsorbed amount) at room temperature on ZEO-1 (black) and USY (red). The blue star indicates the adsorption on as-made ZEO-1, which reveals little external surface adsorption, and the dashed red line indicates reported values of **FAU** (*17*). (**B**) Catalytic performance of several zeolite catalysts in the conversion of heavy oils. Si/Al ratios are indicated at top.

OSDA molecules are located in three distinct, crystallographically independent positions (figs. S13 and S14), which are within the three supercages with four windows of 16×16 MR, 16×12 MR, or 12×12 MR, respectively, with 0.394, 0.499, and 0.495 occupancies. The total number of OSDA molecules obtained from the Rietveld refinement was 44.42 per unit cell, which is in good agreement with the composition found by chemical analysis, $[Si_{583.71}Al_{40.29}O_{1248}]|(C_{19}H_{36}P)_{45.77}(H_2O)_{63.16}$ (table S2).

ZEO-1 has 21 distinct T (tetrahedral) atoms and 43 distinct oxygen atoms, which places it among the most complex fully connected zeolites solved to date. Only zeolites **IMF**, **TUN**, **UOV**, ***STO**, PST-20 (no code assigned so far) (*11*), and ***SFV** (*12*) with 24, 24, 25, 28, 29, and 99 distinct T sites, respectively, are more complex than ZEO-1. It also has a very large unit cell volume of 46,900 $Å^3$, which is only surpassed by the higher members of the isoreticular embedded **RHO** family, **RHO**-G4 (ZSM-25, **MWF**) (*8*) and **RHO**-G5 (PST-20) (*11*). More complex members of this family of small-pore zeolites with even larger volumes have not been solved so far (*13*).

Topological analysis indicates that three different types of tile occur at the channel intersections (Fig. 2, A to C). These three tiles are the supercages with four windows of 16MR and/or 12MR, the ones containing 16MR being larger than the 12MR supercages in zeolites **FAU** (*t-fau*) (Fig. 2D), **EMT** (*t-wou*), **SBS** (*t-ucs* and *t-znf*), and **SBT** (*t-sbt*) (fig. S16). The channel system of ZEO-1 is 3D with interconnected (16+12)×(16+12)×(16+12) MR (Fig. 2, E and F), meaning that the 3D 16MR channel system is interconnected to a second 3D 12MR channel system, and the crystallographic pore apertures of the 16MR and 12MR are 10.62×9.41/10.54×9.64 Å and 7.24×6.60/7.18×5.48 Å, respectively (fig. S17).

Upon calcination in air to 600°C, the organic part of the OSDA within the ZEO-1 channel could be removed while leaving behind phosphorus oxides (fig. S3C) and the framework intact, as proved with PXRD (fig. S8). The phosphorus species can be completely washed out with water (fig. S3D). ZEO-1 is stable up to at least 1000°C (fig. S8). Calcined ZEO-1 exhibited type Ia N_2 and Ar adsorption isotherms (figs. S10 and S11), from which extraordinarily large specific surface area values of 939 and 1037 m^2/g, respectively, were determined with the Brunauer-Emmett-Teller (BET) method. From the Ar adsorption isotherm, the mean pore sizes of ZEO-1 were determined to be 6.7, 7.9, and 11.5 Å (fig. S12), which matched well with the crystallographic values (fig. S17). The Horvath-Kawazoe pore-size distribution (fig. S12) showed a maximum at 8.4 Å, indicating a substantially larger pore size than the reported values for the stable large-pore faujasite (pore entrance ~ 6.8 Å) and the recently reported PST-32 (**SBS**, 6.7 Å) and PST-2 (**SBS**/**SBT** intergrowth, 6.4 Å), all measured by means of Ar adsorption with the same formalism (*9*). Further, two of the supercages in ZEO-1 (those containing 16MR apertures) (Fig. 2, A and B) are larger than the supercages in **FAU**, **SBS**, and **SBT**, whereas the third one (containing only 12MR apertures) (Fig. 2C) is similar in size. ^{29}Si and ^{27}Al nuclear magnetic resonance (figs. S5 and S6) show that ZEO-1 contains framework Al that provides acidity in the calcined state, as determined by the NH_3 adsorption and temperature-programmed desorption (fig. S21) and *n*-decane cracking (fig. S22).

As expected, ZEO-1 allows the entrance of molecules larger than FAU does, as shown by the adsorption of large dye molecules (fig. S20). Although both ZEO-1 and USY readily adsorb methylene blue, ZEO-1 adsorbs >40 times more of the larger Nile red than USY does at saturation (Fig. 3A). With regard to FCC, which converts heavy-oil fractions into lighter, more valuable products (*1*), improved catalysts could increase the yield of fuels and commodities. Currently, zeolite Y (**FAU**) with 3D 12 MR channels and large supercages is the main active component in FCC catalysts (*14*).

Because phosphorus species can have an impact on the performance of zeolite catalysts (*15*), we performed FCC tests on ZEO-1 with and without washing out the occluded P species and on three commercial zeolite catalysts (USY, **MFI**, and Beta). These tests showed a beneficial effect of the occluded phosphorus because the unwashed ZEO-1 has the highest heavy-oil cracking conversion ratio (Fig. 3B and tables S12 and S13) and the highest selectivity to the gasoline and diesel fuel fractions among all the catalysts. The washed catalyst shows a performance closer to that of the best USY catalyst tested but with a larger selectivity to LPG. This change in selectivity from gasoline to LPG upon washing ZEO-1 is likely due to a highly branched gasoline fraction that is prone to cracking in this zeolite but not in

the less acidic P-containing unwashed sample (fig. S21) (*16*). These preliminary laboratory results show that ZEO-1 can compete with the highly optimized USY and also the recently reported PST-32 and PST-2 (*9*) and thus has a high potential as a component of cracking catalysts.

REFERENCES AND NOTES

1. Reza Sadeghbeigi, *Fluid Catalytic Cracking Handbook* (Gulf Publishing, ed. 2, 2000).
2. Ch. Baerlocher, L. B. McCusker, Database of Zeolite Structures (2017); www.iza-structure.org/databases.
3. L. Burel, N. Kasian, A. Tuel, *Angew. Chem. Int. Ed.* **53**, 1360–1363 (2014).
4. H. Xu *et al*., *Angew. Chem. Int. Ed.* **53**, 1355–1359 (2014).
5. Y. Chen *et al*., *Chemistry* **25**, 3219–3223 (2019).
6. Y. Wang *et al*., *Inorg. Chem. Front.* **4**, 1654–1659 (2017).
7. R. Fricke, H.-L. Zubowa, M. Richter, H. Kosslick, *C. R. Chim.* **8**, 549–559 (2005).
8. The structure commission of the International Zeolite Association (SC-IZA) assigns three-letter Zeolite Framework Type codes to approved zeolite structures. Interrupted framework codes are preceded by a hyphen (*2*).
9. H. Lee *et al*., *Science* **373**, 104–107 (2021).
10. G. M. Sheldrick, *Acta Crystallogr. A Found. Adv.* **71**, 3–8 (2015).
11. P. Guo *et al*., *Nature* **524**, 74–78 (2015).
12. C. Baerlocher, T. Weber, L. B. McCusker, L. Palatinus, S. I. Zones, *Science* **333**, 1134–1137 (2011).
13. J. Shin *et al*., *Angew. Chem. Int. Ed.* **55**, 4928–4932 (2016).
14. X. Li, C. Li, J. Zhang, C. Yang, H. Shan, *J. Nat. Gas Chem.* **16**, 92–99 (2007).
15. H. E. van der Bij, B. M. Weckhuysen, *Chem. Soc. Rev.* **44**, 7406–7428 (2015).
16. W. Vermeiren, J.-P. Gilson, *Top. Catal.* **52**, 1131–1161 (2009).
17. I. Pellejero *et al*., *Ind. Eng. Chem. Res.* **46**, 2335–2341 (2007).

ACKNOWLEDGMENTS

We are indebted to M. J. de la Mata (SIdI-UAM) for her expedited help in collecting the solid-state nuclear magnetic resonance spectra and to the ALBA staff for collaboration in collecting the SPXRD data at the Spanish ALBA synchrotron beamline BL04 (MSPD). **Funding:** We acknowledge financial support from the National Natural Science Foundation of China (grants 21601004, 21776312, and 22078364); the Natural Science Foundation of the Higher Education Institutions of Anhui Province, China (grant KJ2020A0585); and MCIN/AEI/10.13039/501100011033, Spain (project PID2019-105479RB-I00). The cRED data was collected at the Electron Microscopy Center (EMC), Department of Materials and Environmental Chemistry (MMK) in Stockholm University, with the support of the Swedish Research Council (grant 1444205) and the Knut and Alice Wallenberg Foundation (KAW) (grant 2012-0112) through the 3DEM-NATUR project. Use of the Advanced Photon Source at Argonne National Laboratory was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, under contract DE-AC02-06CH11357. W.F. gratefully acknowledges support from the US Department of Energy, Office of Science, Basic Energy Sciences, Materials Sciences and Engineering Division, under award DE-SC0019170. **Author contributions:** F.-J.C. designed the project. J.L., X.C., M.A.C., and F.-J.C. supervised the work. Q.-F.L., S.Z., and F.-J.C. carried out the synthesis work. J.L. solved the structure. Z.R.G. and C.L. analyzed the topology. J.C., Z.L., and X.C. carried out the FCC reaction. X.L., W.F., C.L., Z.R.G., and M.A.C. performed the physicochemical characterization. Z.R.G. and M.A.C. performed the large molecule analysis. Z.R.G., C.L., J.L., and M.A.C. prepared the draft. All the authors discussed the results and revised the manuscript. **Competing interests:** Q.-F.L., Z.R.G., C.L., J.L., and F.-J.C. have filed a patent on zeolite ZEO-1 (China patent application no. 202011346698.4). Z.R.G. and J.L. are affiliated with the company holding the rights on that patent. **Data and materials availability:** All data are available in the main text or the supplementary materials. Crystallographic parameters for the structure of as-made and calcined ZEO-1 refined against SPXRD and calcined ZEO-1 refined against cRED data are archived at the Cambridge Crystallographic Data Center (CCDC) (www.ccdc.cam.ac.uk) under reference nos. CCDC 2112774 to 2112776.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abk3258
Materials and Methods
Figs. S1 to S22
Tables S1 to S14
References (*18–28*)
Movie S1
Data Files S1 to S3

6 July 2021; resubmitted 21 July 2021
Accepted 28 October 2021
10.1126/science.abk3258

ULTRAFAST MAGNETISM

Terahertz light–driven coupling of antiferromagnetic spins to lattice

Evgeny A. Mashkovich[1,2]*, Kirill A. Grishunin[1], Roman M. Dubrovin[3], Anatoly K. Zvezdin[4,5], Roman V. Pisarev[3], Alexey V. Kimel[1]

Understanding spin-lattice coupling represents a key challenge in modern condensed matter physics, with crucial importance and implications for ultrafast and two-dimensional magnetism. The efficiency of angular momentum and energy transfer between spins and the lattice imposes fundamental speed limits on the ability to control spins in spintronics, magnonics, and magnetic data storage. We report on an efficient nonlinear mechanism of spin-lattice coupling driven by terahertz light pulses. A nearly single-cycle terahertz pulse resonantly interacts with a coherent magnonic state in the antiferromagnet cobalt difluoride (CoF_2) and excites the Raman-active terahertz phonon. The results reveal the distinctive functionality of antiferromagnets that allows ultrafast spin-lattice coupling using light.

Understanding how to control the spin-lattice interaction is a cornerstone for several hot topics of contemporary magnetic research, including ultrafast magnetism (*1*) and phononic control of magnetism (*2–6*), two-dimensional (2D) magnetism (*7*), magnonics (*8–10*), and spintronics (*11*). Although this interaction is well understood in the vicinity of thermodynamic equilibrium, large-amplitude oscillations of atoms result in an essentially anharmonic lattice dynamics with an increasingly important role of higher-order terms in the atomic free energy (*12*). For instance, driving coherent lattice vibrations into an anharmonic regime can promote energy transfer between different, otherwise noninteracting phononic modes via fully coherent phonon-phonon interactions (*12*). Similarly, one can expect a nonlinear mechanism of light-driven phonon-magnon coupling (*13–15*) and even anticipate nontrivial ultrafast phenomena associated with the physics of the Einstein–de Haas effect (*16*). Although several nonlinear mechanisms of phononic control of magnetism have been demonstrated theoretically (*3*) and experimentally (*4–6*, *17*), alongside intense research interest devoted to exploring terahertz (THz) magnonics, understanding the nonlinear mechanism of energy transfer from THz magnons to THz phonons remains outstanding.

Antiferromagnets represent an appealing playground for the search for new channels of spin-lattice coupling in the THz regime. Their spin structure can be modeled in the simplest case by two antiparallel sublattices with the net magnetizations $\mathbf{M}_1$ and $\mathbf{M}_2$, $|\mathbf{M}_1| = |\mathbf{M}_2|$. Alternatively, for describing the magnetic order, it is convenient to introduce the net magnetization vector $\mathbf{M} = \mathbf{M}_1 + \mathbf{M}_2$ and the Néel (antiferromagnetic) vector $\mathbf{L} = \mathbf{M}_1 - \mathbf{M}_2$. Typically, the frequencies of spin resonances in antiferromagnets are close to those of THz optical phonons. The objective of this study is therefore to use intense, nearly single-cycle THz pulses (*18*) to drive coherent spin oscillations (*9*) and promote interactions between otherwise noninteracting magnonic and phononic modes.

To demonstrate light-driven spin-lattice coupling, we selected a cobalt difluoride (CoF_2) single-crystal plate with the tetragonal rutile crystal structure. Below the Néel temperature $T_N = 39$ K, CoF_2 is a collinear antiferromagnet with a strong piezomagnetic effect (*19*). In a unit cell, the spins of Co^{2+} ions at the cell's center are antiparallel to those at the cell's corners (*20*). These two types of ions form the two antiferromagnetic sublattices with the net

[1]Institute for Molecules and Materials, Radboud University, 6525 AJ Nijmegen, Netherlands. [2]Institute of Physics II, University of Cologne, D-50937 Cologne, Germany. [3]Ioffe Institute, Russian Academy of Sciences, St. Petersburg 194021, Russia. [4]Prokhorov General Physics Institute, Russian Academy of Sciences, Moscow 119991, Russia. [5]Moscow Institute of Physics and Technology, Dolgoprudnyi 141700, Russia.
*Corresponding author. Email: mashkovich@ph2.uni-koeln.de

magnetizations $\mathbf{M}_1$ and $\mathbf{M}_2$, respectively. The magnetizations are aligned along the crystallographic fourfold optical z axis, which is the "easy axis" of magnetic anisotropy (see Fig. 1). The frequency of antiferromagnetic resonance is centered at Ω_m = 1.14 THz at T = 6 K, whereas the nearest phonon mode of the B_{1g} symmetry is at Ω_{ph} = 1.94 THz (*21*, *22*). Strong piezomagnetic properties of CoF_2 imply that atomic and spin dynamics can, in principle, be coupled (*4*, *19*), but because the energies of the phonon and the magnon are substantially different, the coupling is inefficient. In this study, we reveal that a THz photon can fill the magnon-phonon energy gap and thus induce their efficient coupling. More specifically, we show that a nearly single-cycle THz pulse centered at ~1 THz with a bandwidth in excess of $2\Omega_m - \Omega_{ph}$ is able to prepare a coherent magnonic state and subsequently interact with this coherent state by promoting an energy transfer from the coherent magnon to the B_{1g} phonon.

To trace this energy transfer, we used a pump-probe technique to optically detect the coherent phonons and magnons. The atomic motion of the Raman-active B_{1g} phonon mode dynamically breaks the equivalence between the crystallographic x and y axes and thus induces linear birefringence for the light propagating along the optical z axis. Moreover, upon introducing a generalized phonon coordinate $Q_{x^2-y^2}$ corresponding to the B_{1g} phonon and the z component of the Néel vector L_z, it can be shown that if the phonon $Q_{x^2-y^2}$ induces strain σ_{xy}, then the product $L_z\sigma_{xy}$ and the z component of the magnetization M_z transform equivalently under the symmetry operations allowed by the crystallographic 4/mmm point group of CoF_2 (the phonon selection rules) (*23*). This relation can also be attributed to the piezomagnetic effect (*19*), that is, at temperatures below the Néel temperature, coherent atomic motion can also induce magnetic circular birefringence and result in the magneto-optical Faraday effect for light propagating along the z axis. The polarization rotation must be proportional to $L_z\sigma_{xy}$ and therefore should disappear above the Néel temperature. The magnon mode at the frequency of the antiferromagnetic resonance can also be detected optically via magnetic circular birefringence, that is, the Faraday effect, and also via magnetic linear birefringence (the magnon selection rules) (*23*).

The typical THz pump–induced transients for the probe's polarization rotation at different temperatures are shown in Fig. 2A. Polarization rotation with an amplitude of 150 millidegrees (mdeg) is observed below the Néel point and reveals a strong temperature dependence. When normalized to the thickness of the crystal, the signal reaches 3 deg/cm for a THz field on the order of MV/cm, which is comparable with the values reported for similar measurements on NiO (*24*). Figure 2B shows the Fourier transform of the time trace obtained at T = 30 K. It is seen that the THz-pump pulse excites two resonances centered at 1.05 and 1.94 THz. Figure 2C shows that the temperature dependencies of the resonance frequencies in the vicinity of the Néel point are in a perfect agreement with the behaviors expected for the magnon and the B_{1g} phonon in CoF_2 (*25*, *26*). Whereas the magnon mode softens near the Néel point, the phonon frequency does not show any noticeable change in this temperature range. Figure 2D shows the dependencies of the amplitude of the magnon

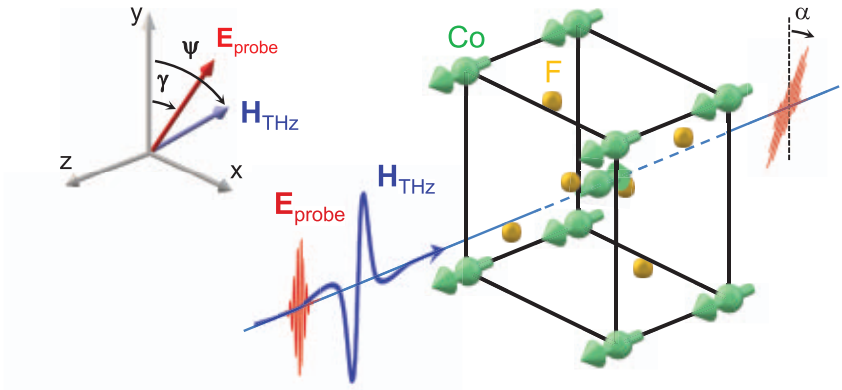


Fig. 1. Geometry of the experiment and THz pumping of antiferromagnetic CoF_2. The spins of Co^{2+} ions are antiferromagnetically aligned along the z axis. The linearly polarized THz pump and near-infrared probe beams propagate approximately collinearly along the z axis and spatially overlap on the sample. By varying the time delay t_{det} between the infrared probe and the THz-pump pulses, we measure pump-induced ultrafast dynamics. The angle ψ between the THz pulse magnetic field $\mathbf{H}_{THz}$ and the y axis is tuned in the range of $\pm\pi/2$ using two wire-grid polarizers. The polarization of the probe pulse forms an angle γ with the y axis and is controlled by a half-wave plate.

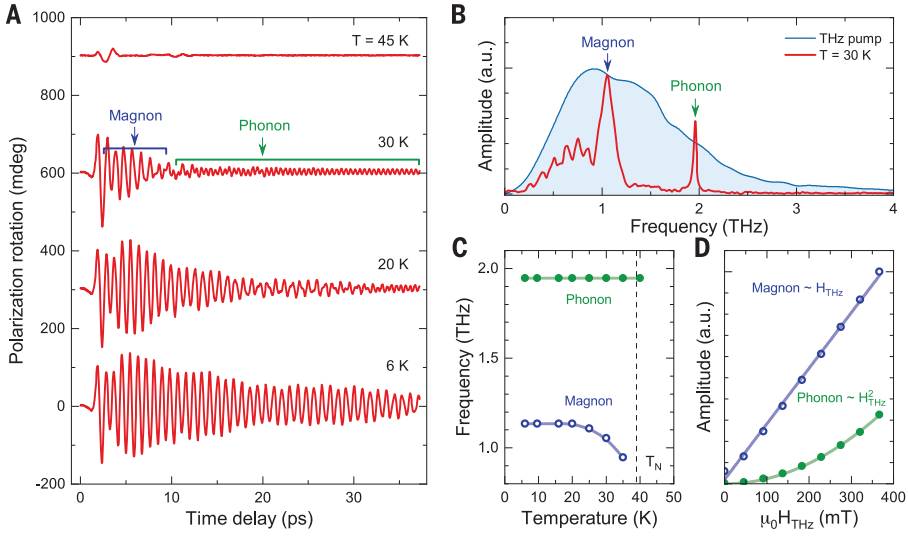


Fig. 2. THz excitation of magnons and phonons. (**A**) THz pump–induced rotation of the probe's polarization measured at different temperatures for the case of the horizontal orientation of the probe electric field (γ = 90°) and the THz magnetic field (ψ = 90°). (**B**) Fourier spectrum of the waveform from (A) measured at T = 30 K. The spectrum of the incident THz pulse is shown by the light blue area. (**C**) The frequencies of the magnon mode (blue open circles) and the phonon mode (green filled circles) deduced from the Fourier spectrum as a function of temperature. The solid lines are guides for the eye. The dashed line marks the Néel temperature. (**D**) The Fourier spectral amplitudes of the magnon mode (blue open circles) and the phonon mode (green filled circles) at T = 6 K as functions of the THz magnetic field strength $H_{THz} = |\mathbf{H}_{THz}|$. The solid lines represent linear and quadratic fits. The corresponding waveforms are plotted in figs. S8 and S9. a.u., arbitrary units.

and the phonon resonances on the THz magnetic field strength. The magnon amplitude is a linear function of the field, which is typical for the conventional mechanism of excitation of the antiferromagnetic resonance by a magnetic field via the Zeeman torque (*8*, *24*). By contrast, the phonon amplitude scales quadratically with the field strength similarly to that in (*27*), thus clearly evidencing the nonlinear mechanism of the excitation. A spectrogram of the probe polarization rotation at T = 31 K is plotted in fig. S10.

To describe the dynamics of the Néel vector **L**, one can represent it as a sum of the stationary $\mathbf{L}_0$ and the THz-induced **l** parts: $\mathbf{L} = \mathbf{L}_0 + \mathbf{l}$, where $\mathbf{L}_0 >> \mathbf{l}$. From the Lagrange-Euler equations (*28*, *29*), one finds that the dynamics of **l** is described by differential equations for damped harmonic oscillators (the magnon selection rules) (*23*). The latter gives

$$l_x(t) \sim \tilde{H}(\Omega_m)\cos(\Omega_m t + \xi)e^{-t/\tau_m}\cos\psi$$
$$l_y(t) \sim \tilde{H}(\Omega_m)\cos(\Omega_m t + \xi)e^{-t/\tau_m}\sin\psi \quad (1)$$

where $\tilde{H}(\Omega_m)$ is the spectral amplitude of the THz magnetic field at Ω_m, τ_m is the magnon damping time, and ξ is the phase. Hence, any orientation of the magnetic field in the xy plane excites oscillations of **l**. Note that the equations of motion for **l**, which can be derived either from the Lagrange-Euler or directly from the Landau-Lifshitz-Gilbert (*30*) equations, show that these oscillations are launched in the plane orthogonal to the THz pulse magnetic field $\mathbf{H}_{THz}$.

In the case of atomic motion at the B_{1g} phonon frequency with the generalized phononic coordinate $Q_{x^2-y^2}$, one can write a conventional equation of motion for the harmonic oscillator (*12*)

$$\frac{d^2Q_{x^2-y^2}}{dt^2} + \frac{2}{\tau_{ph}}\frac{dQ_{x^2-y^2}}{dt} + \Omega_{ph}^2 Q_{x^2-y^2} = T_{x^2-y^2} \quad (2)$$

where the first term in the left part corresponds to acceleration and the second term accounts for damping where τ_{ph} is the phonon damping time. The third term corresponds to the restoring force, and $T_{x^2-y^2}(t)$ is the driving torque. Obviously, the equation must be invariant under all symmetry operations of the 4/mmm point group of CoF_2, that is, the torque $T_{x^2-y^2}(t)$ must transform equivalently to $Q_{x^2-y^2}$ (see table S1). Taking into account the irreducible representations for this point group (*31*) and focusing on terms of the second order with respect to the THz electric and magnetic fields, the following four contributions to $T_{x^2-y^2}(t)$ are allowed:

$$T_{x^2-y^2}(t) = C_1\left[E_x^2(t) - E_y^2(t)\right] + C_2\left[H_x^2(t) - H_y^2(t)\right] + C_3[l_x(t)H_y(t) - l_y(t)H_x(t)] + C_4\left[l_x^2(t) - l_y^2(t)\right] \quad (3)$$

where C_1, C_2, C_3, and C_4 are phenomenological coefficients; E_i and H_i are the i-component of the THz magnetic and electric fields, correspondingly. The first and second terms correspond to electric and magnetic dipole mechanisms of two-photon excitation of the B_{1g} phonon via a virtual state. A similar mechanism was described in (*32*, *33*). In contrast to the two-photon excitation of phonons via an intermediate phononic state, as proposed in (*34*), the third term in Eq. 3 describes two-photon excitation of phonons via an intermediate magnonic state. The fourth term shows that the phonon can also be excited because of the anharmonicity of magnons. However, one may argue that in the case of forced oscillations of the antiferromagnetic Néel vector by a THz magnetic field, the effects of the third and the fourth terms must be similar and can hardly be distinguished.

The temperature dependencies of the detected signal at the phonon and magnon frequencies are shown in Fig. 3A. The substantially low (~30 times) phononic oscillations above the Néel point is a result of the dominating role of the magneto-optical Faraday effect in the detection of the phonon. Because the phonon-induced Faraday rotation of the probe pulse is proportional to the $L_z\sigma_{xy}$ product, the detection efficiency becomes zero above the Néel temperature. Intriguingly, the detected phononic amplitude stays nearly constant in the range from 6 to 25 K and even has a maximum at 30 K, where the equality $\Omega_{ph} = 2\Omega_m$ holds. By contrast, the magnon frequency Ω_m smoothly and monotonically decreases from 1.14 THz at T = 6 K to zero at the Néel temperature, and this frequency change is accompanied by a broadening of the magnon spectral line $\Delta\Omega_m$ (see fig. S4). Such changes affect the efficiencies of phononic excitations in different ways because of the mechanisms that correspond to different terms in Eq. 3. To elucidate this, we modeled temperature dependencies of the mechanisms that correspond to the first, third, and fourth terms in Eq. 3. In particular, our model explored how temperature affects the strength of the oscillations induced at the frequency of the phonon (the phonon selection rules) (*23*). In the modeling, we also took into account that the phonon and magnon are detected optically and that the sensitivity of the detection is practically proportional to the Néel vector. Comparing the simulated dependencies with that observed experimentally clearly emphasizes the dominating role of the magnon-mediated mechanism in the excitation of the phonon (see Fig. 3B).

To demonstrate that the nonlinear excitation of phonons is mediated by a coherent magnonic state, we performed double-pump experiments using two identical THz pulses separated by a time delay t_{exc} with a peak magnetic field of 100 mT. For different t_{exc}, we measured the probe's polarization rotation as a function of time delay between the first

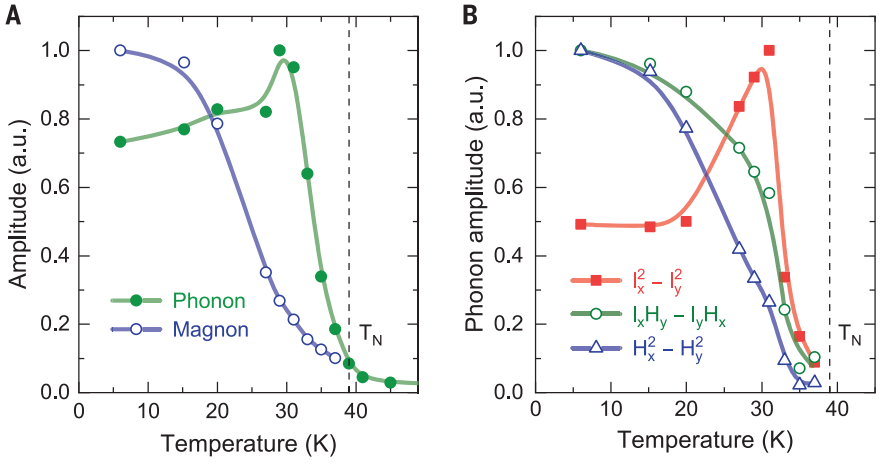


Fig. 3. Phonon and magnon excitation. (**A**) Temperature dependencies of the Fourier amplitudes of the magnon mode (blue open circles) and the phonon mode (green filled circles). The solid lines are guides to the eye. The dashed line marks the Néel temperature. The corresponding waveforms are given in fig. S8. (**B**) Estimated phononic amplitude excited via the virtual state (blue open triangles; first and second terms in Eq. 3), real magnonic state (green filled circles; the third term in Eq. 3), and anharmonic magnon (red filled squares; the fourth term in Eq. 3). The dashed line marks the Néel temperature.

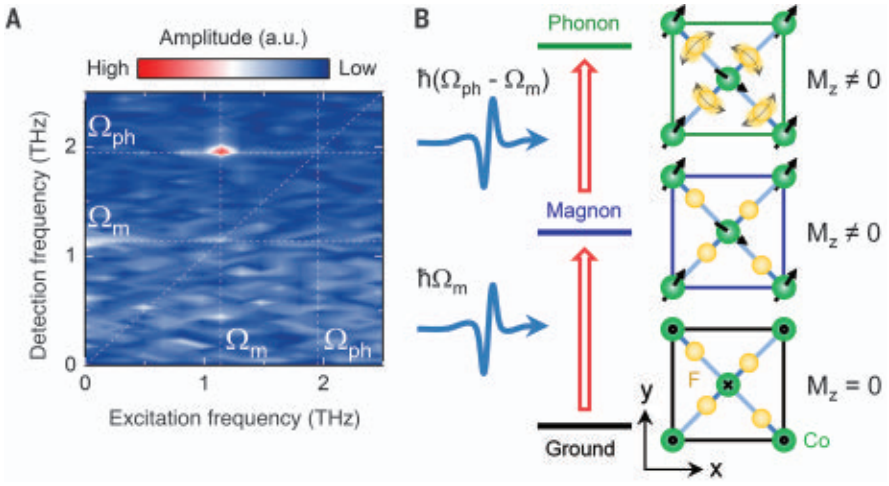


Fig. 4. THz light–driven coupling of lattice to spins. (**A**) 2D Fourier spectrum of the nonlinear amplitude $\tilde{\alpha}_{NL}(f_{exc}, f_{det})$. The measurements were performed at T = 10 K for horizontal orientations of the probe electric field (γ = 90°) and the THz magnetic field (ψ = 90°). (**B**) A pictorial of the magnon-mediated excitation of the B_{1g} phonon by a THz magnetic field. The unperturbed antiferromagnetic state of the CoF_2 unit cell is shown at the bottom. A THz photon resonantly populates the coherent magnonic state at the frequency Ω_m, thus creating an intermediate state. Another THz photon at the frequency of $\Omega_{ph} - \Omega_m$ interacts with this intermediate state and coherently excites the B_{1g} phonon. $\hbar$, reduced Planck's constant.

THz pump and the probe pulse t_{det}. Using the same method of data acquisition as explained in (*35*), we measured the dynamics triggered by both THz-pump pulses $\alpha_{oo}(\tau)$; by the first pump only, that is, when the second THz pump is blocked $\alpha_{oc}(\tau)$; and by the second THz pump only (the first THz pump is blocked) $\alpha_{co}(\tau)$. Here, τ is the vector $\tau = (t_{exc}; t_{det})$. We deduce the intrinsically nonlinear response of the medium by subsequently calculating the difference between the signals in the time-domain:

$$\alpha_{NL}(\tau) = \alpha_{oo}(\tau) - \alpha_{oc}(\tau) - \alpha_{co}(\tau) \quad (4)$$

where $\alpha_{NL}(\tau)$ is nonzero only if the first pump changes the spin susceptibility to the THz magnetic field. A Fourier transform of the extracted signals $\alpha_{NL}(\tau)$ allows us to plot the spectrum as a 2D graph $\tilde{\alpha}_{NL}(f_{exc}; f_{det})$ (see Fig. 4A). Here, f_{det} and f_{exc} are the Fourier frequencies of the corresponding time delays t_{det} and t_{exc}, respectively. The spectrum $\tilde{\alpha}_{NL}(f_{exc}; f_{det})$ clearly reveals one maximum at $f_{det} = \Omega_{ph}$ and $f_{exc} = \Omega_m$, revealing that the excitation of the B_{1g} phonon via a nonlinear THz light–driven mechanism is only possible when the magnon is excited. This means that the macroscopic coherent magnonic state with the frequency Ω_m, resonantly generated by the broadband THz pulse, has an essentially different susceptibility to the THz magnetic field compared with that associated with its unperturbed state. More particularly, the interaction of the second THz photon at the frequency $\Omega_{ph} - \Omega_m$ with the coherent magnonic state is able to excite the B_{1g} phonon at the frequency Ω_{ph} (see Fig. 4B).

The ultrafast coherent transfer of spin energy to the lattice driven by THz light pulses opens up possibilities for the fields of nonlinear phononics, ultrafast magnetism, THz magnonics, and antiferromagnetic spintronics. The dynamic coupling of the lattice to the spins is crucial for ultrafast control of magnetism and phase engineering through transient changes of magnetic and structural states. Exactly the same light-driven spin-lattice coupling is allowed in other antiferromagnetic fluorides that have the same point group (MnF_2, FeF_2, and others), but the effect must be more general and can be extended to other materials. Because a coherent magnonic state in any ferro-, ferri-, and even antiferromagnetic material can induce dynamic magnetization, the excitation of this coherent state with a magnetic field at the frequency matching the gap between the magnon and the phonon might even lead to nontrivial ultrafast phenomena associated with the physics of the Einstein–de Haas effect. This mechanism is especially appealing for ultrafast coherent control of materials with complex charge and spin ordering, for example, in multiferroics and 2D magnets.

REFERENCES AND NOTES

1. C. Dornes *et al.*, *Nature* **565**, 209–212 (2019).
2. D. M. Juraschek, P. Narang, N. A. Spaldin, *Phys. Rev. Res.* **2**, 043035 (2020).
3. D. M. Juraschek, D. S. Wang, P. Narang, *Phys. Rev. B* **103**, 094407 (2021).
4. A. S. Disa *et al.*, *Nat. Phys.* **16**, 937–941 (2020).
5. A. Stupakiewicz *et al.*, *Nat. Phys.* **17**, 489–492 (2021).
6. D. Afanasiev *et al.*, *Nat. Mater.* **20**, 607–611 (2021).
7. K. S. Burch, D. Mandrus, J.-G. Park, *Nature* **563**, 47–52 (2018).
8. E. A. Mashkovich *et al.*, *Phys. Rev. Lett.* **123**, 157202 (2019).
9. S. Baierl *et al.*, *Nat. Photonics* **10**, 715–718 (2016).
10. S. Schlauderer *et al.*, *Nature* **569**, 383–387 (2019).
11. J. Walowski, M. Münzenberg, *J. Appl. Phys.* **120**, 140901 (2016).
12. M. Först *et al.*, *Nat. Phys.* **7**, 854–856 (2011).
13. S. Streib, N. Vidal-Silva, K. Shen, G. E. W. Bauer, *Phys. Rev. B* **99**, 184442 (2019).
14. S. Streib, H. Keshtgar, G. E. W. Bauer, *Phys. Rev. Lett.* **121**, 027202 (2018).
15. P. Maldonado, Y. O. Kvashnin, *Phys. Rev. B* **100**, 014430 (2019).
16. A. Rückriegel, S. Streib, G. E. W. Bauer, R. A. Duine, *Phys. Rev. B* **101**, 104402 (2020).
17. T. F. Nova *et al.*, *Nat. Phys.* **13**, 132–136 (2017).
18. H. Hirori, A. Doi, F. Blanchard, K. Tanaka, *Appl. Phys. Lett.* **98**, 091106 (2011).
19. A. S. Borovik-Romanov, *J. Exp. Theor. Phys.* **38**, 1088–1098 (1960).
20. R. A. Erickson, *Phys. Rev.* **90**, 779–785 (1953).
21. P. L. Richards, *J. Appl. Phys.* **35**, 850–851 (1964).
22. R. M. Macfarlane, S. Ushioda, *Solid State Commun.* **8**, 1081–1083 (1970).
23. See supplementary materials.
24. T. Kampfrath *et al.*, *Nat. Photonics* **5**, 31–34 (2011).
25. M. G. Cottam, D. J. Lockwood, *Low Temp. Phys.* **45**, 78–91 (2019).
26. E. Meloche, M. G. Cottam, D. J. Lockwood, *Low Temp. Phys.* **40**, 134–145 (2014).
27. M. Kozina *et al.*, *Nat. Phys.* **15**, 387–392 (2019).
28. A. F. Andreev, V. I. Marchenko, *Sov. Phys. Usp.* **23**, 21–34 (1980).
29. A. K. Zvezdin, A. A. Mukhin, *Bull. Lebedev Phys. Inst.* **12**, 10 (1981).
30. A. V. Kimel, A. M. Kalashnikova, A. Pogrebna, A. K. Zvezdin, *Phys. Rep.* **852**, 1–46 (2020).
31. A. S. Borovik-Romanov, *Lectures on Low-Temperature Magnetism: Magnetic Symmetry of Antiferromagnets* (Tsifrovichok, Moscow, 2010).
32. S. Maehrlein, A. Paarmann, M. Wolf, T. Kampfrath, *Phys. Rev. Lett.* **119**, 127402 (2017).
33. K. Imasaka *et al.*, *Phys. Rev. B* **98**, 054303 (2018).
34. G. Khalsa, N. A. Benedek, J. Moses, *Phys. Rev. X* **11**, 021067 (2021).
35. J. Lu *et al.*, *Top. Curr. Chem.* **376**, 6 (2018).
36. E. A. Mashkovich *et al.*, Terahertz light–driven coupling of antiferromagnetic spins to lattice, Dataset, Zenodo (2021); https://doi.org/10.5281/zenodo.5557846.

ACKNOWLEDGMENTS

We thank S. Semin and C. Berkhout for technical support. We acknowledge fruitful discussions with M. A. Prosnikov and C. Davies. **Funding:** This research was funded by the Netherlands Organization for Scientific Research (NWO). The contribution of E.A.M. was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation), project number 277146847 - CRC 1238. R.M.D. and R.V.P. acknowledge financial support from the Russian Foundation for Basic Research according to project no. 19-02-00457. The theoretical contribution of A.K.Z. was supported by the Russian Science Foundation (the Project N 17-12-01333). **Author contributions:** Conceptualization: E.A.M., R.V.P., A.V.K.; Methodology: E.A.M., K.A.G., A.K.Z.; Investigation: E.A.M., K.A.G., R.M.D.; Resources: R.M.D., R.V.P.; Supervision: A.V.K.; Writing – original draft: E.A.M.; Writing – review and editing: E.A.M., K.A.G., R.M.D., A.K.Z., R.V.P., A.V.K. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are deposited at Zenodo (*36*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abk1121
Materials and Methods
Supplementary Text
Figs. S1 to S10
Table S1
References (*37*–*46*)

24 June 2021; accepted 11 November 2021
10.1126/science.abk1121

BIOCATALYSIS

Stereodivergent atom-transfer radical cyclization by engineered cytochromes P450

Qi Zhou[1], Michael Chin[1]†, Yue Fu[2]†, Peng Liu[2], Yang Yang[1,3]*

Naturally occurring enzymes can be a source of unnatural reactivity that can be molded by directed evolution to generate efficient biocatalysts with valuable activities. Owing to the lack of exploitable stereocontrol elements in synthetic systems, steering the absolute and relative stereochemistry of free-radical processes is notoriously difficult in asymmetric catalysis. Inspired by the innate redox properties of first-row transition-metal cofactors, we repurposed cytochromes P450 to catalyze stereoselective atom-transfer radical cyclization. A set of metalloenzymes was engineered to impose substantial stereocontrol over the radical addition step and the halogen rebound step in these unnatural processes, allowing enantio- and diastereodivergent radical catalysis. This evolvable metalloenzyme platform represents a promising solution to tame fleeting radical intermediates for asymmetric catalysis.

As nature's privileged catalysts, enzymes are well known for their ability to exert exquisite control over the stereochemical outcome of chemical reactions (*1*). Over the past three decades, the advent of directed evolution (*2*, *3*) has enabled the rapid development of customized enzymes to generate excellent activity and stereoselectivity, which are often complementary to those of traditional small-molecule catalysts (*4*). However, until recently, the catalytic repertoire of enzymes has been mostly limited to reactions found in nature, posing constraints on the types of products available from enzyme catalysis. To merge the excellent tunability and stereocontrol of biocatalysts with the synthetic versatility of abiotic systems, the discovery and development of new-to-nature enzymatic activity are widely recognized as preeminent objectives at the interface of modern biocatalysis and organic synthesis (*5*, *6*).

Our group initiated a research program to repurpose naturally occurring metalloenzymes to catalyze unnatural radical reactions in a stereocontrolled fashion. Owing to the reactive nature of radical intermediates and the lack of synthetically achievable stereoinduction strategies, imposing enantio- and diastereocontrol over free-radical–mediated bond forming processes has remained a daunting challenge in asymmetric catalysis (*7*, *8*). On the other hand, natural enzymes (*9*), such as radical *S*-adenosylmethionine enzymes (*10*), facilitate radical reactions with excellent chemo-, regio-, and stereoselectivity. However, most natural radical enzymes display a narrow substrate scope. We propose that repurposing widely used, biotechnologically important enzymes to catalyze new-to-nature radical reactions may allow a diverse array of easily available substrates to be converted with operational simplicity. In this respect, recent notable contributions from the Hyster and Zhao laboratories (*11*) demonstrated the use of visible light to unveil excited-state activities of ketoreductases (*12*) and ene reductases (*13*–*18*), allowing for a range of stereoselective hydrogen-transfer transformations to be developed.

Small-molecule complexes of first-row transition metals are reported to catalyze unselective radical reactions via ground-state single-electron transfer between the metal center and the substrate, and we envisioned a powerful strategy to control the stereoselectivity of these radical reactions by engaging enzymes bearing a redox-active metallocofactor in these unnatural pro-

[1]Department of Chemistry and Biochemistry, University of California Santa Barbara, Santa Barbara, CA 93106, USA. [2]Department of Chemistry, University of Pittsburgh, Pittsburgh, PA 15260, USA. [3]Biomolecular Science and Engineering Program, University of California Santa Barbara, Santa Barbara, CA 93106, USA.
*Corresponding author. Email: yang@chem.ucsb.edu
†These authors contributed equally to this work.

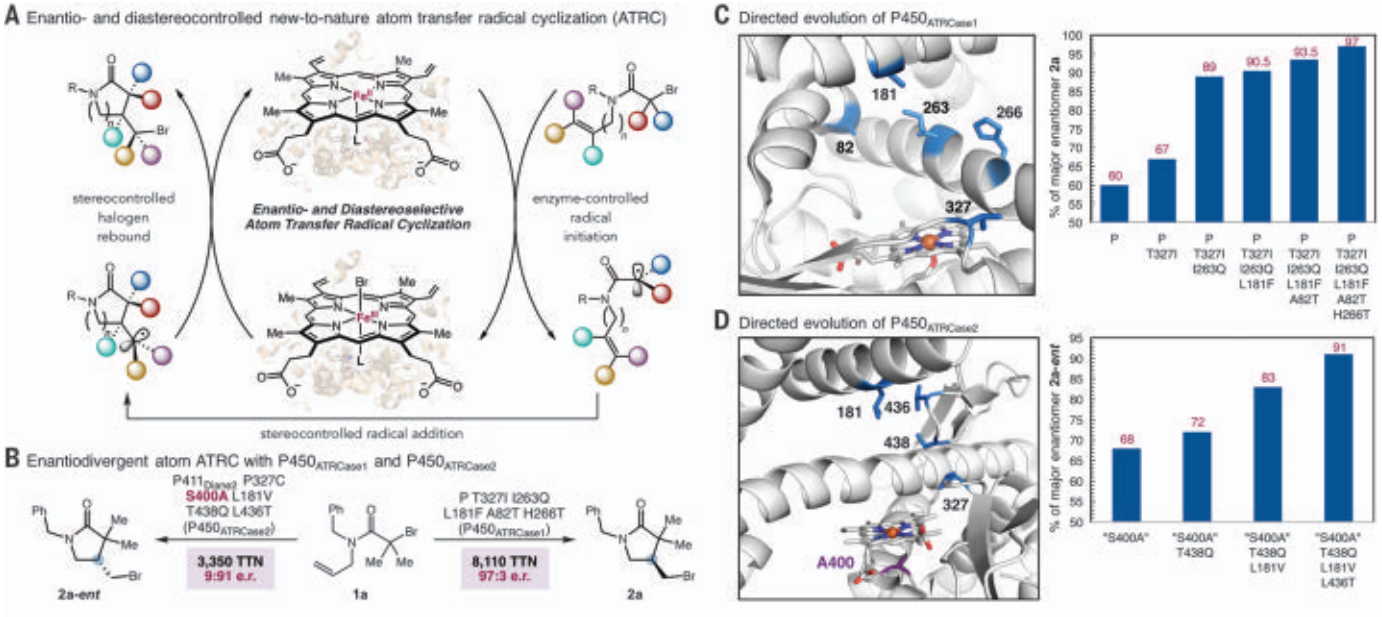


Fig. 1. New-to-nature biocatalytic ATRC. (**A**) Enantio- and diastereocontrolled ATRC. R, alkyl; Me, methyl; L, Fe-coordinating amino acid residue. (**B**) Enantiodivergent ATRC using $P450_{ATRCase1}$ and $P450_{ATRCase2}$. Ph, phenyl. (**C**) Directed evolution of $P450_{ATRCase1}$. The illustration at left was created on the basis of the crystal structure of a closely related P450 variant [Protein Data Bank identifier (PDB ID): 4H23]. "P" denotes P411-CIS T438S. (**D**) Directed evolution of $P450_{ATRCase2}$. The illustration at left was created on the basis of the crystal structure of a closely related P450 variant (PDB ID: 5UCW). "S400A" denotes $P411_{Diane2}$ with P327C and S400A. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; Y, Tyr.

cesses. By fine-tuning the interaction between the reactive radical intermediate and the metalloprotein scaffold by means of directed evolution, these metalloenzyme catalysts could be engineered to achieve levels of enantio- and diastereocontrol that are unmatched by small-molecule catalysts. With this strategy, a diverse array of unnatural radical metalloenzymes could potentially be developed, as metalloredox couples that span a wide range of potentials—including Fe(II)/Fe(III) (*19*, *20*), Co(I)/Co(II) (*21*), and Cu(I)/Cu(II) (*22*)—are readily available in a myriad of native metalloenzymes. In particular, inspired by polymer chemists' research in employing heme-based systems to generate radical species for polymerization (*23*, *24*), we saw numerous opportunities for leveraging heme proteins as a general platform for asymmetric radical transformations. Previous important work from Arnold, Fasan, and others demonstrated that heme proteins can be repurposed to catalyze unnatural carbene and nitrene transfers analogous to the natural oxene transfer observed in heme oxygenases (*25*–*30*). Thus, the highly promiscuous and evolvable nature of heme proteins makes them particularly promising for the development of stereocontrolled metalloredox radical chemistry.

We envisioned a family of unnatural biocatalytic reactions—namely, stereoselective atom-transfer radical cyclization (ATRC) (*31*, *32*)—to be advanced with metalloenzymes. In our proposed ATRC catalyzed by Fe-dependent enzymes (Fig. 1A), the metalloprotein catalyst in its ferrous state undergoes single-electron transfer with an organic halide, producing a transient radical intermediate in the active site. Rapid addition of the nascent radical species to an unsaturated system would afford a product radical, which would subsequently react with the ferric halide to produce the product and regenerate the metalloenzyme catalyst. The envisioned ATRC simultaneously installs an alkyl group and a halogen atom across the C=C double bond of an olefin substrate, thereby generating valuable products with added stereochemical complexity (*31*). Despite the extensive study and widespread utility of ATRC reactions, the development of general methods for stereocontrolled ATRC presents substantial hurdles for small-molecule catalysts (*7*, *33*, *34*). Because it is difficult to maintain a tight association between the chiral catalyst and the radical intermediate, general catalytic strategies to exert enantiocontrol over the C–C bond forming radical addition step remain unavailable. Imposing catalyst-controlled diastereoselectivity in the subsequent halogen-rebound step is similarly challenging. Herein, we describe the development of metalloprotein catalysts that provide excellent enantio- and diastereocontrol for this daunting problem in asymmetric catalysis. Notably, our protein engineering efforts allow for development of a toolbox of stereocomplementary catalysts, granting enantio- and diastereodivergent access to ATRC products.

In the present study, we evaluated both heme (P450s, globins, and cytochromes *c*) and nonheme Fe-dependent proteins for the targeted transformation (table S1). ATRC of substrate **1a** was selected as the model reaction, because $FeCl_2$ was known to facilitate these processes (*35*) and the corresponding enantioenriched lactams derived from these cyclization processes are ubiquitous in medicinally relevant scaffold (*36*). The utility of Fe enzymes in our library was evaluated in the form of purified proteins, cell-free lysates, and whole-cell catalysts. Although several metalloproteins were found to promote the desired radical cyclization (table S1), among all of the metalloprotein catalysts we examined, a serine-ligated variant of *Bacillus megaterium* P450 (CYP102A1) (*37*)—known as "P" (*38*)—furnished a measurable enantiomeric excess [60:40 enantiomeric ratio (e.r.)] and the highest activity. Notably, the activity of intact *Escherichia coli* cells harboring P was an order of magnitude higher than that of the purified protein catalyst with identical enantioselectivity (tables S4 and S5). The ability to use whole bacterial cells as catalysts without further manipulation substantially accelerated this protein engineering campaign.

To further improve the enantioselectivity of this radical cyclization process, we next performed directed evolution through iterative cycles of site-saturation mutagenesis (SSM) and screening (Fig. 1, B and C). In each round of engineering, 90 clones were screened. By

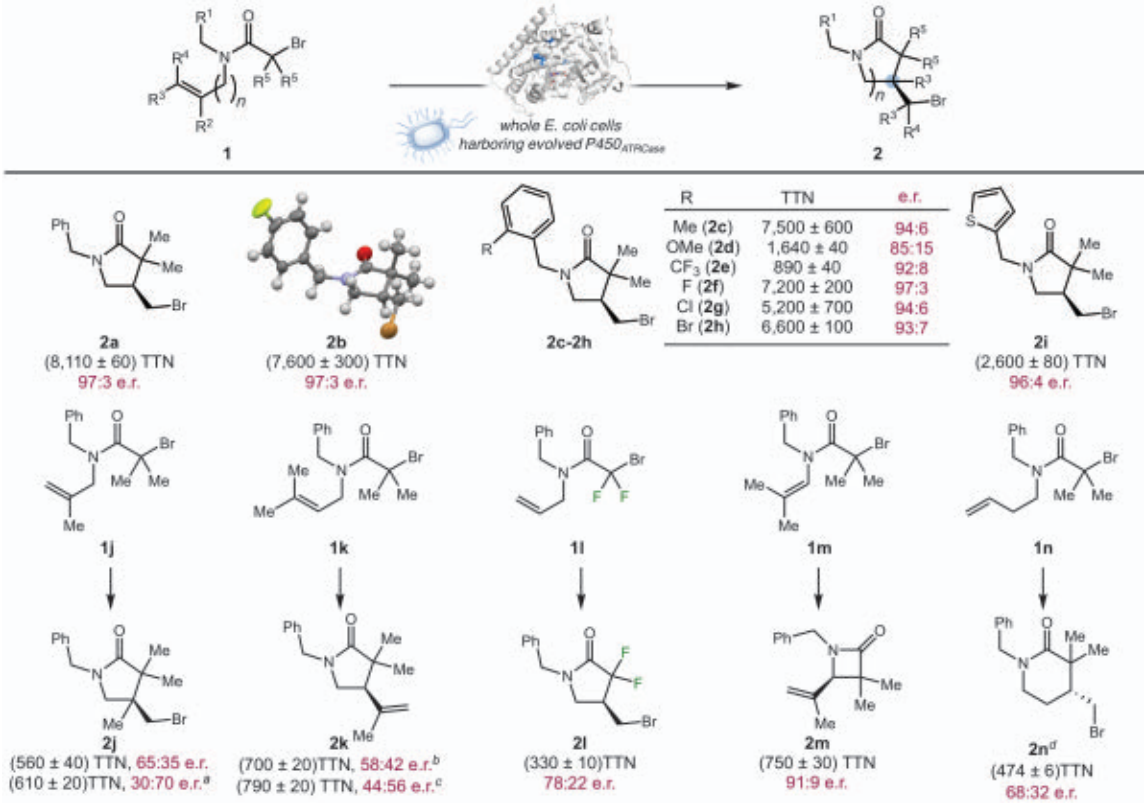


Fig. 2. Substrate scope of new-to-nature enantioselective ATRC. [a]Variant P′ was used. [b]P with T327I and I263Q was used. [c]$P450_{ATRCase1}$ (P with T327I, I263Q, L181F, A82T, and H266T) was used. [d]$P450_{ATRCase2}$ with A330K was used. Whole-cell reactions were carried out at OD_{600} (optical density at 600 nm) = 5 to 30. The variation of e.r. values is ≤1%. See the supplementary materials for details.

targeting active-site residues proximal to the heme cofactor—including residues 327, 263, 181, 82, and 266—through five rounds of protein engineering, the e.r. of the desired product (**2a**) could be rapidly enhanced. With our final metalloenzyme variant, $P450_{ATRCase1}$ (P with T327I, I263Q, L181F, A82T, and H266T), the desired product formed with excellent yield and enantioselectivity [89% yield, a total turnover number (TTN) of 8110, and 97:3 e.r.]. By further decreasing the catalyst loading, a TTN of 20,000 was achieved without affecting the enantioselectivity.

During the course of this study, we also found that another serine-ligated P450 variant ($P411_{Diane2}$) that we previously engineered for enantioselective C–H amidation (*39*) was also active for this ATRC, with inverted enantiopreference (Fig. 1B). $P411_{Diane2}$ was further evolved via recursive SSM and screening (Fig. 1D). An axial ligand effect was investigated in this unnatural metalloredox radical chemistry. Notably, the "S400A" variant lacking a coordinating axial ligand (here,"S400A" indicates $P411_{Diane2}$ with the P327C and S400A mutations) facilitated the unnatural radical reaction with a 1.4-fold improvement in TTN relative to the S400 variant, without affecting the enantioselectivity (from 1160 TTN and 32:68 e.r. to 1590 TTN and 32:68 e.r.; table S7). By targeting other active-site residues, including 327, 181, 438, and 436, we arrived at $P450_{ATRCase2}$ ($P411_{Diane2}$ with P327C, S400A, L181V, T438Q, and L436T) with substantially enhanced activity and enantioselectivity (3350 TTN, 91:9 e.r.).

We next surveyed the substrate scope of the ATRC process (Fig. 2). A range of functional groups on the nitrogen substituent was readily tolerated, providing the corresponding γ-lactam in excellent TTN and enantioselectivity (**2a** to **2i**). Chloro- (**2g**) and bromo- (**2h**) substituents were also compatible, providing a valuable functional group handle for further derivatization. Substrates bearing a heterocycle, such as thiophene (**2i**), were also transformed with excellent enantiocontrol. Substituted olefins (**1j** and **1k**) as well as α,α-difluorinated (**1l**) substrates could also be converted into the desired γ-lactam products, including **2j** bearing contiguous quaternary-quaternary stereocenters. The absolute configuration of **2b** was ascertained by x-ray crystallography. Using these engineered metalloenzyme catalysts, β- and δ-lactams (**2m** and **2n**) readily formed in an enantioselective fashion, indicating the potential of this biocatalytic platform to access a diverse array of enantioenriched lactam products. Products **2k** and **2m** were derived from dehydrohalogenation of the ATRC product.

Given the success of engineered metalloenzymes in exerting stereocontrol over the C–C bond forming step, we questioned whether we could further leverage this metalloredox biocatalytic platform to control the relative stereochemistry in the C–Br bond forming halogen rebound event (Fig. 3A). To this end, we re-examined all of the P450 variants we accumulated in this protein engineering effort. We identified two hits, $P450_{ATRCase}$ gen-4 (P with T327I, I263Q, L181F, and A82T) and P′ (P with L181V, L437F, and S438Q), giving rise to promising levels of opposite diastereopreferences. Through iterative rounds of SSM

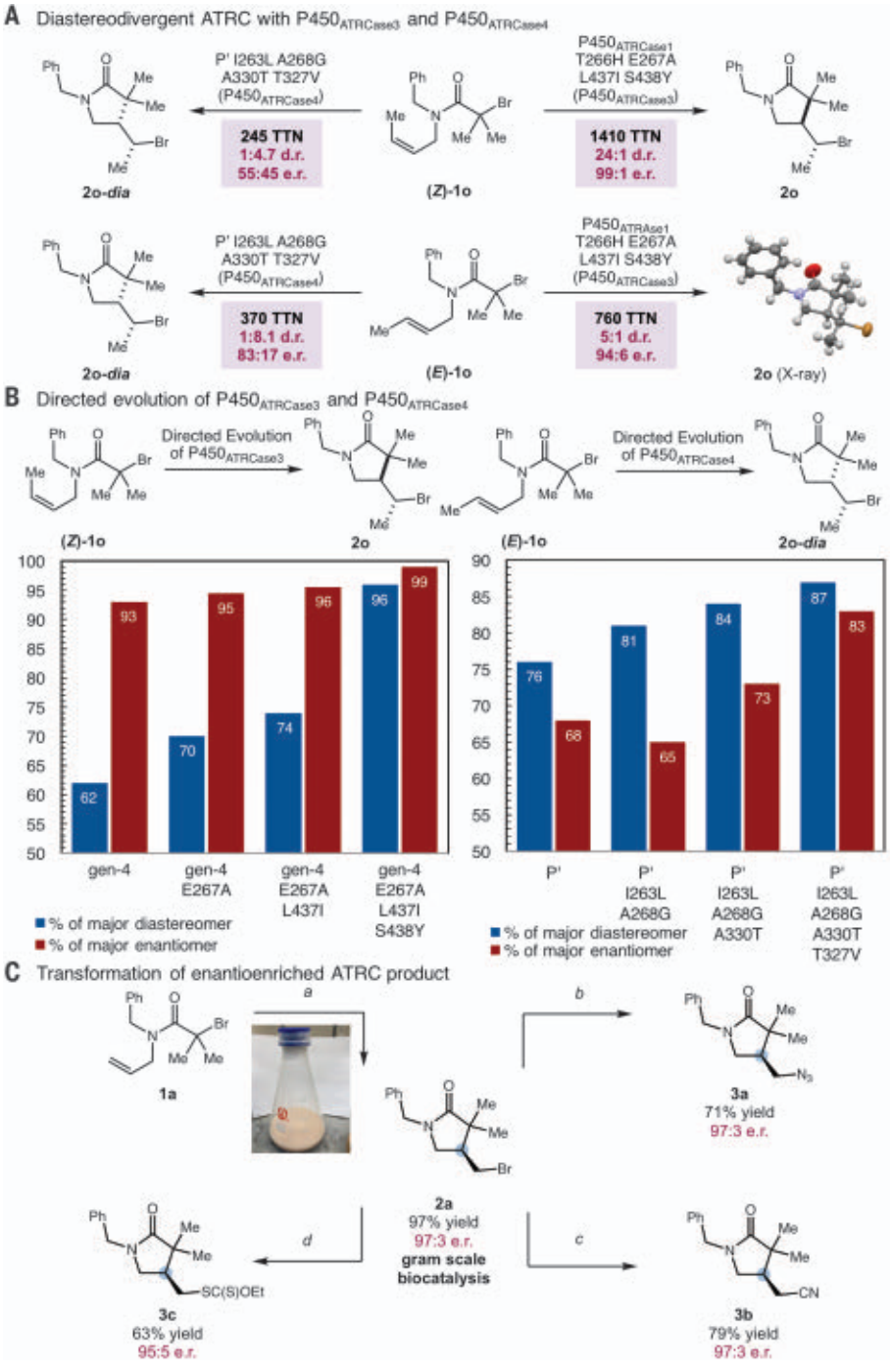


Fig. 3. Further development and application of new-to-nature ATRC. (**A**) Diastereodivergent new-to-nature ATRC with $P450_{ATRCase3}$ and $P450_{ATRCase4}$. (**B**) Directed evolution of $P450_{ATRCase3}$ and $P450_{ATRCase4}$. (**C**) Transformation of enantioenriched ATRC product. Conditions: *a*. Whole-cell biotransformation using *E. coli* cells resuspended in M9-N buffer (pH = 7.4) at room temperature (RT) for 12 hours, 97% yield, 97:3 e.r.; *b*. NaN_3, NaI, dimethylformamide (DMF)/H_2O, 60°C, 16 hours, 79% yield, 97:3 e.r.; *c*. NaCN, NaI, DMF/H_2O, 60°C, 16 hours, 71% yield, 97:3 e.r.; and *d*. KSC(S)OEt, acetone, RT, 12 hours, 63% yield, 95:5 e.r..

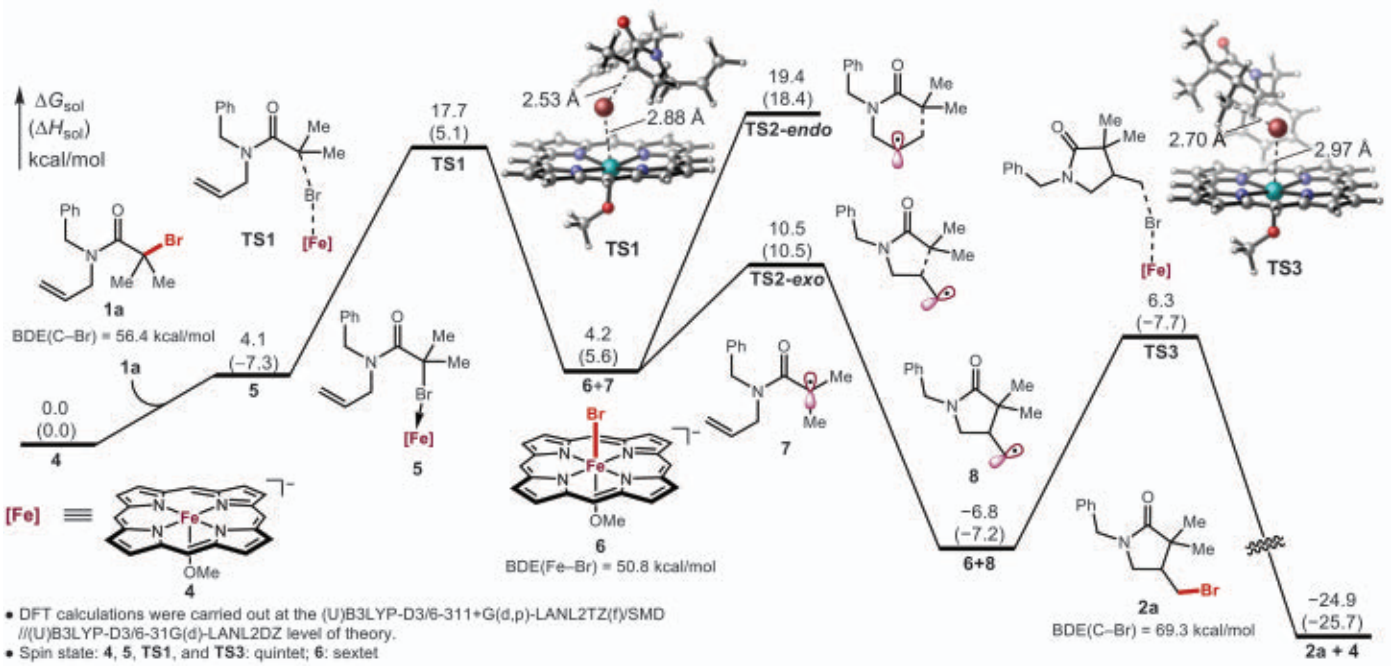


Fig. 4. Reaction energy profile of the current biocatalytic atom-transfer radical addition using a model system for the axial serine-ligated Fe-porphyrin catalyst. A methoxy-bound Fe porphyrin was used as the model system for serine-ligated P450 (*39*). ΔG_{sol}, Gibbs free energy; ΔH_{sol}, enthalpy; BDE, bond dissociation enthalpy.

and screening, we engineered an orthogonal set of highly selective biocatalysts, including $P450_{ATRCase3}$ and $P450_{ATRCase4}$, allowing for the diastereodivergent synthesis of either **2o** or **2o-*dia*** (Fig. 3B). For comparison, the use of the traditional tris(2-pyridylmethyl)amine copper(I) catalyst for ATRC (*40*) produced **2o** and **2o-*dia*** with low diastereoselectivity [1:1.6 and 2.1:1 diastereomeric ratio from (*E*)- and (*Z*)-**1o**, respectively; see the supplementary materials for details]. Moreover, with $P450_{ATRCase3}$ and $P450_{ATRCase4}$, (*E*)- and (*Z*)-**1o** provided the same major diastereomer (see the supplementary materials). Thus, starting from easily accessible mixtures of (*E*)- and (*Z*)-**1o**, the diastereoconvergent biocatalytic transformation with $P450_{ATRCase3}$ or $P450_{ATRCase4}$ delivered the corresponding γ-lactam product in good diastereoselectivity. The relative stereochemistry of **2o** was determined by x-ray diffraction analysis (see the supplementary materials for details).

To further demonstrate the synthetic utility of engineered radical metalloenzymes, we performed this whole-cell biotransformation on a gram scale (Fig. 3C). Excellent yield and enantioselectivity were observed, showcasing the practicality of this new-to-nature biocatalytic reaction. Furthermore, the presence of a bromine functional group in enantioenriched ATRC products allowed for a range of diversification reactions to be conveniently carried out. S_N2-type (i.e., second order) nucleophilic substitution with a range of nucleophiles generated synthetically versatile azide (**3a**), cyanide (**3b**), and xanthate (**3c**) products in good yields while maintaining stereochemical purity. Combined with the enantioselective biocatalytic ATRC, these derivatization reactions allowed a range of formal enantioselective carbofunctionalization reactions, including carboazidation (**3a**), carbocyanation (**3b**), and carboxanthation (**3c**), to be achieved.

Consistent with our initial hypothesis, our radical clock experiments with hemin as the catalyst led to ring-opening products (see the supplementary materials for details), indicating that radical intermediates are involved in this process. To gain further insight into the reaction mechanism, we performed density functional theory (DFT) calculations using a model system (*39*) for the axial serine-ligated P450 catalyst (Fig. 4; see the supplementary materials for details). DFT calculations showed that the Fe porphyrin catalyst remains high-spin throughout the catalytic cycle in this biocatalytic ATRC process. This finding is not consistent with the previously studied native oxene transfer (*41*) and analogous nitrene transfer (*39*) chemistry of P450 enzymes, wherein spin crossover is involved. With the model system, the radical initiation step (**TS1**) has a relatively low activation barrier ($\Delta G^{\ddagger}$) of 17.7 kcal/mol. Considering that the enzyme environment may further facilitate this process by promoting substrate binding to form complex **5**, this Fe-catalyzed radical initiation is expected to be kinetically facile. The electron-rich nature of the Fe center allows for the facile single-electron reduction of the substrate, as evidenced by the substantial electron transfer (0.44 e^-) from **4** to **1a** in **TS1**. After the selective 5-*exo*-trig cyclization (**TS2-exo**) to form the primary carbon radical **8**, the bromine rebound step (**TS3**) is highly exergonic, with a low activation barrier of 13.1 kcal/mol. The fast trapping of the carbon radical via bromine atom transfer renders the C–C bond formation step irreversible and enables kinetic control of reaction stereochemistry. The bromine rebound reactivity is promoted by the conversion of a weaker Fe–Br bond in the ferric bromide species (**6**) to a stronger primary $C(sp^3)$–Br bond in **2a**. Therefore, the ability of the Fe porphyrin system to promote both radical initiation and bromine rebound steps makes it an effective ATRC biocatalyst.

REFERENCES AND NOTES

1. U. T. Bornscheuer *et al.*, *Nature* **485**, 185–194 (2012).
2. J. D. Bloom, F. H. Arnold, *Proc. Natl. Acad. Sci. U.S.A.* **106** (suppl. 1), 9995–10000 (2009).
3. U. T. Bornscheuer, B. Hauer, K. E. Jaeger, U. Schwaneberg, *Angew. Chem. Int. Ed.* **58**, 36–40 (2019).
4. M. T. Reetz, *Angew. Chem. Int. Ed.* **50**, 138–174 (2011).
5. H. Renata, Z. J. Wang, F. H. Arnold, *Angew. Chem. Int. Ed.* **54**, 3351–3367 (2015).
6. K. Chen, F. H. Arnold, *Nat. Catal.* **3**, 203–213 (2020).
7. M. P. Sibi, S. Manyem, J. Zimmerman, *Chem. Rev.* **103**, 3263–3296 (2003).
8. R. S. J. Proctor, A. C. Colgan, R. J. Phipps, *Nat. Chem.* **12**, 990–1004 (2020).
9. C. M. Jäger, A. K. Croft, *ChemBioEng Rev.* **5**, 143–162 (2018).
10. J. B. Broderick, B. R. Duffus, K. S. Duschene, E. M. Shepard, *Chem. Rev.* **114**, 4229–4317 (2014).

11. B. A. Sandoval, T. K. Hyster, *Curr. Opin. Chem. Biol.* **55**, 45–51 (2020).
12. M. A. Emmanuel, N. R. Greenberg, D. G. Oblinsky, T. K. Hyster, *Nature* **540**, 414–417 (2016).
13. B. A. Sandoval, A. J. Meichan, T. K. Hyster, *J. Am. Chem. Soc.* **139**, 11313–11316 (2017).
14. K. F. Biegasiewicz *et al*., *Science* **364**, 1166–1169 (2019).
15. P. D. Clayman, T. K. Hyster, *J. Am. Chem. Soc.* **142**, 15673–15677 (2020).
16. M. J. Black *et al*., *Nat. Chem.* **12**, 71–75 (2020).
17. X. Huang *et al*., *Nature* **584**, 69–74 (2020).
18. C. G. Page *et al*., *J. Am. Chem. Soc.* **143**, 97–102 (2021).
19. E. G. Kovaleva, J. D. Lipscomb, *Nat. Chem. Biol.* **4**, 186–193 (2008).
20. I. G. Denisov, T. M. Makris, S. G. Sligar, I. Schlichting, *Chem. Rev.* **105**, 2253–2277 (2005).
21. K. L. Brown, *Chem. Rev.* **105**, 2075–2149 (2005).
22. E. I. Solomon *et al*., *Chem. Rev.* **114**, 3659–3853 (2014).
23. K. J. Rodriguez *et al*., *ACS Macro Lett.* **7**, 1111–1119 (2018).
24. A. Simakova, M. Mackenzie, S. E. Averick, S. Park, K. Matyjaszewski, *Angew. Chem. Int. Ed.* **52**, 12148–12151 (2013).
25. Y. Yang, F. H. Arnold, *Acc. Chem. Res.* **54**, 1209–1225 (2021).
26. O. F. Brandenberg, R. Fasan, F. H. Arnold, *Curr. Opin. Biotechnol.* **47**, 102–111 (2017).
27. T. Hayashi *et al*., *Nat. Catal.* **1**, 578–584 (2018).
28. H. M. Key, P. Dydio, D. S. Clark, J. F. Hartwig, *Nature* **534**, 534–537 (2016).
29. P. Dydio *et al*., *Science* **354**, 102–106 (2016).
30. D. M. Carminati, R. Fasan, *ACS Catal.* **9**, 9683–9697 (2019).
31. W. T. Eckenhoff, T. Pintauer, *Catal. Rev., Sci. Eng.* **52**, 1–59 (2010).
32. A. J. Clark, *Eur. J. Org. Chem.* **2016**, 2231–2243 (2016).
33. B. Chen, C. Fang, P. Liu, J. M. Ready, *Angew. Chem. Int. Ed.* **56**, 8780–8784 (2017).
34. D. Yang, S. Gu, Y.-L. Yan, N.-Y. Zhu, K.-K. Cheung, *J. Am. Chem. Soc.* **123**, 8612–8613 (2001).
35. C. K. Tseng, E. G. Teach, R. W. Simons, *Synth. Commun.* **14**, 1027–1031 (1984).
36. E. Vitaku, D. T. Smith, J. T. Njardarson, *J. Med. Chem.* **57**, 10257–10274 (2014).
37. C. J. C. Whitehouse, S. G. Bell, L.-L. Wong, *Chem. Soc. Rev.* **41**, 1218–1260 (2012).
38. J. A. McIntosh *et al*., *Angew. Chem. Int. Ed.* **52**, 9309–9312 (2013).
39. Y. Yang, I. Cho, X. Qi, P. Liu, F. H. Arnold, *Nat. Chem.* **11**, 987–993 (2019).
40. C. Schumacher, J. G. Hernández, C. Bolm, *Angew. Chem. Int. Ed.* **59**, 16357–16360 (2020).
41. S. Shaik, D. Kumar, S. P. de Visser, A. Altun, W. Thiel, *Chem. Rev.* **105**, 2279–2328 (2005).

ACKNOWLEDGMENTS

We are grateful to B. Lipshutz, J. Read de Alaniz, A. Zakarian, and L. Zhang (UCSB) for their advice and generous support. We thank Y. Wang (University of Pittsburgh) for helpful discussions and comments on the manuscript. L. Lu is acknowledged for assistance with radical clock studies. **Funding:** This work is supported by the start-up funds from the University of California Santa Barbara (Y.Y.) and the National Institutes of Health (R35GM128779; P.L.). We acknowledge the BioPACIFIC MIP (NSF Materials Innovation Platform, DMR-1933487) at UCSB for access to instrumentation. Q.Z. is an Otis Williams Postdoctoral Fellow in Bioengineering. Calculations were performed at the Center for Research Computing at the University of Pittsburgh. **Author contributions:** Y.Y. designed the overall research. Y.Y. and Q.Z. performed initial evaluation of metalloenzymes and reaction conditions. Q.Z., M.C., and Y.Y. performed directed evolution of the $P450_{ATRCase}$ metalloenzymes. Q.Z. and M.C. performed substrate scope study. Y.F. carried out the computational studies, with P.L. providing guidance. Y.Y. wrote the manuscript, with input from all other authors. M.C. and Y.F. contributed equally, and their names appear alphabetically in the author list. **Competing interests:** Y.Y., Q.Z., and M.C. are inventors on a patent application (US provisional patent application no. 63/228,562) submitted by the University of California Santa Barbara that covers stereoselective biocatalytic atom-transfer radical addition processes. **Data and materials availability:** All data are available in the main text or the supplementary materials. Solid-state structures of **2b** and **2o** are available from the Cambridge Crystallographic Data Centre under reference numbers CCDC 2087196 and 2087196. Plasmids encoding engineered enzymes are available from Y.Y. under a material transfer agreement with the University of California Santa Barbara.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abk1603
Materials and Methods
Figs. S1 to S6
Tables S1 to S17
References (*42*–*102*)

25 June 2021; accepted 15 October 2021
10.1126/science.abk1603

CARBON NANOTUBES

Semiconductor nanochannels in metallic carbon nanotubes by thermomechanical chirality alteration

Dai-Ming Tang[1]*, Sergey V. Erohin[2], Dmitry G. Kvashnin[2,3], Victor A. Demin[3], Ovidiu Cretu[4], Song Jiang[5], Lili Zhang[5], Peng-Xiang Hou[5], Guohai Chen[6], Don N. Futaba[6], Yongjia Zheng[7], Rong Xiang[7], Xin Zhou[1], Feng-Chun Hsia[1], Naoyuki Kawamoto[4], Masanori Mitome[1], Yoshihiro Nemoto[8], Fumihiko Uesugi[8], Masaki Takeguchi[8], Shigeo Maruyama[7], Hui-Ming Cheng[5,9,10], Yoshio Bando[11,12], Chang Liu[5]*, Pavel B. Sorokin[2,13]*, Dmitri Golberg[1,14]*

Carbon nanotubes have a helical structure wherein the chirality determines whether they are metallic or semiconducting. Using in situ transmission electron microscopy, we applied heating and mechanical strain to alter the local chirality and thereby control the electronic properties of individual single-wall carbon nanotubes. A transition trend toward a larger chiral angle region was observed and explained in terms of orientation-dependent dislocation formation energy. A controlled metal-to-semiconductor transition was realized to create nanotube transistors with a semiconducting nanotube channel covalently bonded between a metallic nanotube source and drain. Additionally, quantum transport at room temperature was demonstrated for the fabricated nanotube transistors with a channel length as short as 2.8 nanometers.

Single-wall carbon nanotubes (SWCNTs) have a one-dimensional helical tubular molecular structure made up of hexagonally bonded sp^2 carbon atoms. Conceptually, a SWCNT could be formed by rolling up a graphene sheet along a so-called chiral vector. The chirality of a SWCNT uniquely determines its atomic geometry and electronic structure, that is, whether it is metallic or semiconducting (*1*). Semiconducting CNTs can be used to make energy-efficient nanotransistors (*2*–*4*) and so are promising for building beyond-silicon microprocessors (*5*, *6*). However, despite progress in selective growth and separation (*7*–*10*), it remains a great challenge to control the chirality of individual CNTs. Molecular junctions between metallic and semiconducting nanotubes can also be used as the basis for nanoscale electronic devices (*11*), and intramolecular CNT junctions have exhibited nonlinear transport characteristics similar to rectifying diodes (*12*, *13*). The junctions formed at randomly occurring defects during the nanotube growth. Mechanical strain can modify the electronic properties, and theoretically, a plastic strain can change CNT chirality (*14*, *15*) to create intramolecular nanotube junctions. Experimentally, modification of the CNT chirality by plastic deformation has been reported (*16*, *17*), but these transformations, as well as the resulting electrical properties, were not controlled but rather were considered random jumps caused by small energy differences between different chiral structures (*18*).

Herein, we report the design and fabrication of a CNT intramolecular transistor where the local chirality is altered by thermomechanical processing in a controlled manner within a transmission electron microscope (TEM).

[1]International Center for Materials Nanoarchitectonics (MANA), National Institute for Materials Science (NIMS), Tsukuba 305-0044, Japan. [2]National University of Science and Technology (MISIS), Moscow 119049, Russian Federation. [3]Emanuel Institute of Biochemical Physics, Moscow 119334, Russian Federation. [4]Research Center for Advanced Measurement and Characterization, National Institute for Materials Science (NIMS), Tsukuba 305-0044, Japan. [5]Shenyang National Laboratory for Materials Science, Institute of Metal Research, Chinese Academy of Sciences, Shenyang 110016, China. [6]CNT-Application Research Center, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba 305-8565, Japan. [7]Department of Mechanical Engineering, The University of Tokyo, Tokyo 113-8656, Japan. [8]Electron Microscopy Analysis Station, National Institute for Materials Science (NIMS), Tsukuba 305-0047, Japan. [9]Shenzhen Geim Graphene Center, Tsinghua-Berkeley Shenzhen Institute, Tsinghua University, Shenzhen 518055, China. [10]Faculty of Materials Science and Engineering/Institute of Technology for Carbon Neutrality, Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen 518055, China. [11]Institute of Molecular Plus, Tianjin University, Tianjin 300072, China. [12]Australian Institute for Innovative Materials, University of Wollongong, North Wollongong NSW 2500, Australia. [13]Moscow Institute of Physics and Technology, Moscow Region 141701, Russian Federation. [14]Centre for Materials Science and School of Chemistry and Physics, Queensland University of Technology (QUT), Brisbane QLD 4000, Australia.
***Corresponding author. Email: tang.daiming@nims.go.jp (D.-M.T.); cliu@imr.ac.cn (C.L.); pbsorokin@misis.ru (P.B.S.); dmitry.golberg@qut.edu.au (D.G.)**

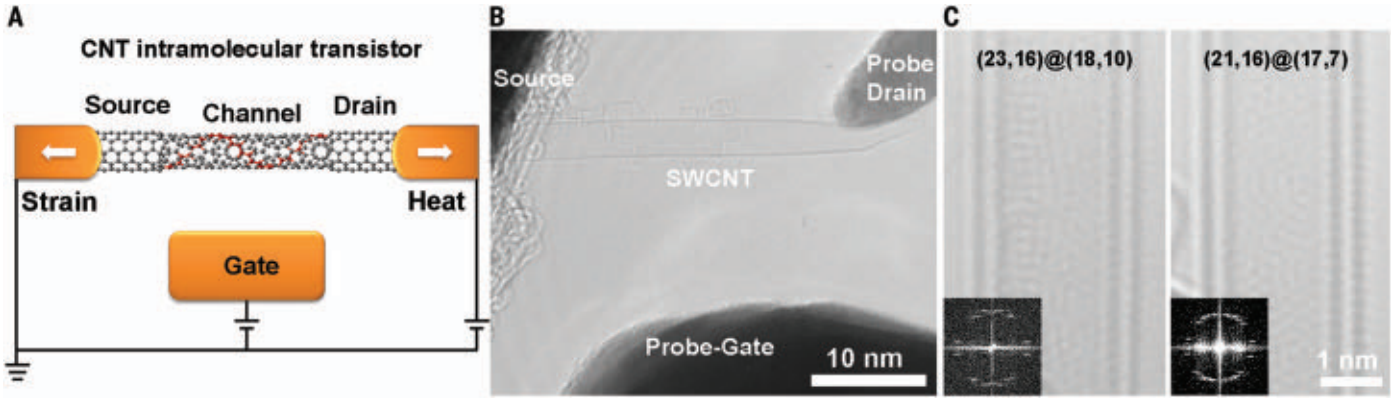


Fig. 1. Fabrication and characterization of CNT intramolecular transistors. (**A**) Schematic of a CNT intramolecular transistor with local chirality altered by mechanical strain and Joule heating. (**B**) TEM image of a SWCNT transistor with a fixed source electrode, a probe as drain electrode, and another probe as gate electrode. (**C**) Lattice-resolved TEM images and corresponding FFT patterns of a DWCNT before and after the chirality transition.

Straining a heated metallic CNT led to a chirality transition of an internal segment that became a semiconductor, thus creating an intramolecular nanotube transistor (Fig. 1A). These transistors had a channel length as short as 2.8 nm and exhibited coherent quantum interference at room temperature.

The in situ TEM method combined nanomanipulation, atomic characterization, and transistor measurements. Step-by-step electrical, thermal, and mechanical stimuli were applied to suspended individual CNTs using two piezo-controlled probes within the TEM. The detailed fabrication process can be found in the methods section of the supplementary materials and fig. S1. In short, nanotubes were attached to the edge of a metal wire or a TEM grid by direct growth or transfer after growth. Individual nanotubes protruding from the metal edge were located and then approached by the probes under TEM monitoring. Plastic deformation of the nanotube was activated and localized to the middle hot segment under resistive (Joule) heating to induce a local chirality change. The localized deformation is related to the one-dimensional thermal transport of the CNTs and good contacts with the electrodes. We used a computer program (fig. S1) to control the strain and voltage to keep a near-transition condition with a temperature high enough (>2000 K) to initiate the plastic deformation and a relatively small strain step (~1%) to localize the deformation region. A bias ranging from 2.0 to 3.0 V was applied to an individual CNT, resulting in a current of tens of microamperes. The total duration of the transformation process was ~10 s. The electron beam current density was kept at a low value (~0.5 pA/cm^2 at the fluorescent TEM screen) to avoid irradiation damage. After each processing cycle, changes in the CNT chirality were determined by electron diffraction patterns and spherical aberration (Cs)–corrected TEM images with atomic resolution. Electrical transport properties were measured in a suspended transistor configuration using a stationary electrode as the source electrode, one probe as the drain electrode, and another probe as the gate electrode (Fig. 1B).

As an example of these changes, TEM images of a double-wall carbon nanotube (DWCNT) before and after a chirality transition show distinct Moiré patterns from overlaid upper- and lower-wall lattices (Fig. 1C). Analysis of corresponding fast Fourier transform (FFT) patterns showed that the chirality indices changed from (23,16)@(18,10) to (21,16)@(17,7). During the transition, local rotation and distortion of the lattice were observed, indicating the presence of dislocations (fig. S2). After the transition, uniform graphitic lattices showed that the channel was largely defect-free. The electrical properties of a SWCNT section stretched between two multiwall carbon nanotube sections were monitored as the length was increased from ~1.1 to 7.7 nm and the resistance increased from ~16.7 to ~22.9 kilohm (fig. S3). Both the initial resistance and the small increase of resistance after elongation indicated that the transport was near the ballistic regime. Although defects such as dislocations were generated during chirality transformations, they were highly mobile at the high temperature caused by Joule heating. After the dislocation gliding, a junction was formed between the deformed and undeformed regions, and the channel was largely defect-free and therefore highly conductive.

The fabrication process of a SWCNT intramolecular transistor is shown in TEM images (Fig. 2, A to C) and a schematic (Fig. 2D). The initial diameter of the SWCNT was ~7.7 nm (Fig. 2A). Because of the large diameter, the nanotube was metallic (Fig. 2, E and F), as confirmed by the gate voltage (V_G) independent conductance and the nearly linear source-drain (SD) current-voltage (I_{SD}-V_{SD}) curve with a conductance of ~3.73 × 10^{-5} S, around half the quantum conductance (~7.75 × 10^{-5} S, or near-ballistic transport at room temperature). Thermomechanical processing was performed under a bias of 3.0 V and a current of ~6.5 μA with a pulse duration of 0.1 ms. After eight stretching cycles, the channel length increased from ~15.2 to ~26.1 nm, and the diameter decreased from ~7.7 to 2.7 nm.

The resulting chirality transitions of the SWCNT turned the stretched region into a semiconductor with an on-off current (I_{ON}/I_{OFF}) ratio of ~4.3 × 10^3 (Fig. 2E) and a nonlinear current-voltage (*I-V*) curve (Fig. 2F). The transfer curve indicated ambipolar transport characteristics, with the electron and hole branches being largely symmetric, which is expected for a molecular junction between a metallic nanotube and an intrinsic semiconducting nanotube with the Fermi level at the middle of the bandgap (*19*). After another three stretching cycles, the tube diameter was further reduced to ~1.7 nm, and the gate voltage required to turn on the current became larger, confirming a widened bandgap inversely proportional to the CNT diameter (*I*).

In addition, reversible transitions between metallic and semiconducting behavior were observed with consecutive cycles (fig. S4). By monitoring the electrical properties as a feedback signal during repeated thermomechanical processing (typically 10 cycles), the desired transition to semiconducting CNTs could be obtained in a controlled manner. The method was used to alter the chirality of SWCNTs grown by chemical vapor deposition (CVD). Figure S5 shows a SWCNT in which the chiral indices changed from (23,5) to (15,4) along with a transition from metallic to semiconducting behavior that was confirmed by a V_G-dependent conductance measurement.

We used this method to fabricate a SWCNT transistor with a diameter of ~0.6 nm and a channel length as short as ~2.8 nm (Fig. 2, G to I). For such a narrow channel under a drive

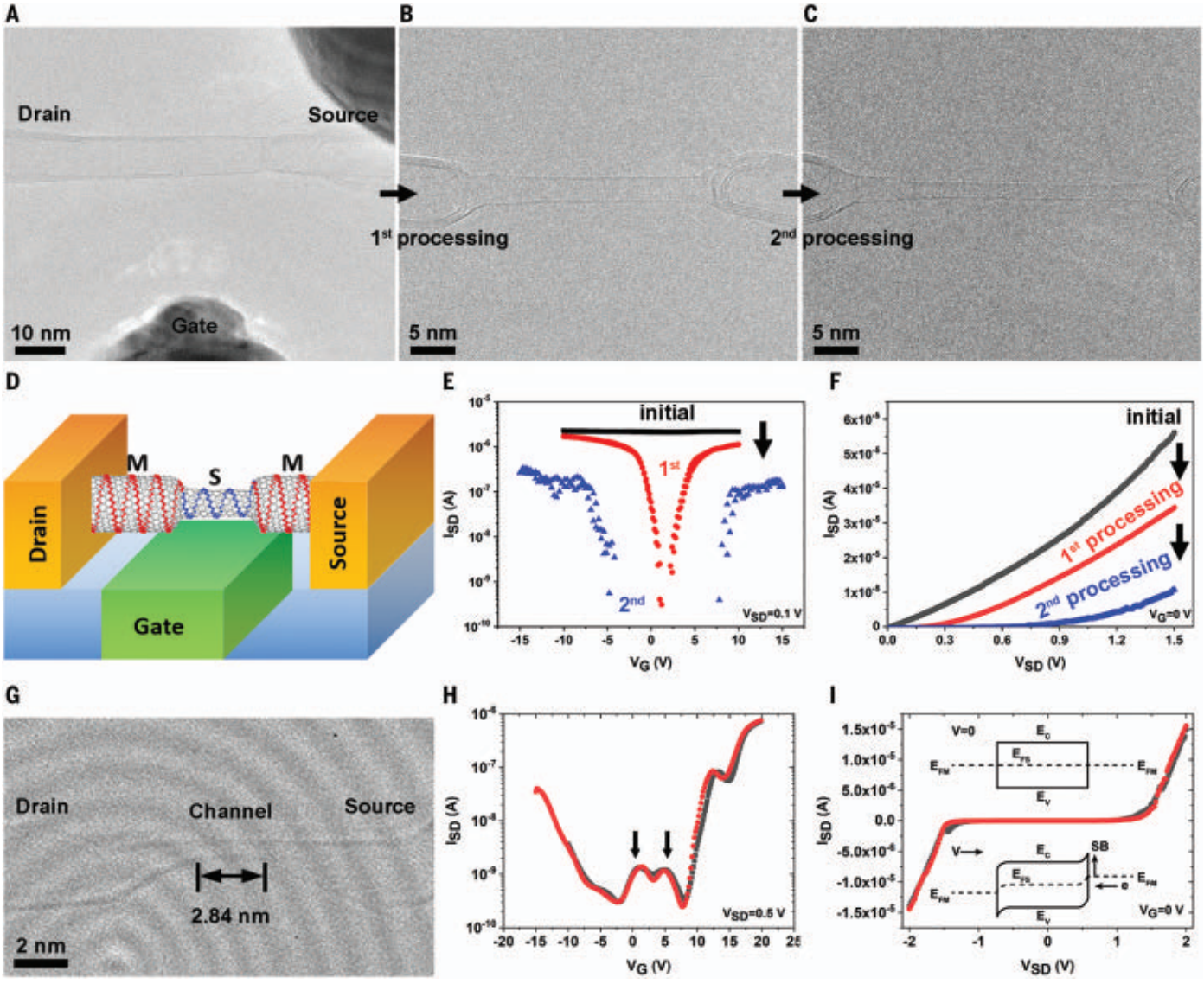


Fig. 2. Metal-to-semiconductor transition for fabricating SWCNT intramolecular transistors. (**A** to **C**) TEM images of a SWCNT through consecutive thermomechanical processing cycles. (**D**) Schematic of a SWCNT intramolecular transistor. M, metal; S, semiconductor. (**E** and **F**) I_{SD}-V_G and I_{SD}-V_{SD} transport curves of the initial tube and after transformations, showing the transition from metallic to semiconducting behavior and the widening of the bandgap. (**G**) TEM image of a SWCNT intramolecular transistor with a channel length of ~2.8 nm. (**H**) I_{SD}-V_G transport measurements of the ultrashort channel transistor. Resonant conductance peaks in the gap are indicated by arrows. (**I**) I_{SD}-V_{SD} characteristic with V_G at 0 V. Insets are schematics of the band structures of the metallic-semiconducting-metallic nanotube junctions. E_{FM}, Fermi level of the metallic nanotube; E_C, conduction band; E_{FS}, Fermi level of the semiconducting nanotube; E_V, valence band; SB, Schottky barrier.

voltage of 0.5 V, the ON current was as high as ~0.74 μA, corresponding to a current density of ~1233 μA/μm with the channel width normalized to the tube diameter. The OFF current was ~0.2 nA, and the I_{ON}/I_{OFF} ratio was ~3000. The subthreshold swing (SS) was ~1.33 V/dec, which is lower than that of previously reported CNT transistors in a suspended configuration (*20*), where for a channel length of ~30 nm, the SS value was ~4.9 V/dec.

The transistor performance is closely dependent on the geometry, interface bonding, and electrical resistance of the source and drain contacts (*21*). Atomically thin graphene contacts were reported to improve electrostatic control of the gate (*3*), and end-contacts were critical to reducing the contact resistance for CNT nanotransistors (*4*). In our work, the semiconducting CNT channel was covalently bonded to the metallic CNT leads, which naturally made good contacts to external electrodes. Such seamless contact with a needlelike geometry has been theoretically predicted to be ideal for nanotube transistors (*22*), where the strong σ bonds enhance the cohesion and the delocalized π electrons result in low charge-carrier scattering at the interface.

The metallic-semiconducting-metallic nanotube junctions form a symmetric structure of double Schottky barriers. A schematic of the band structure of the CNT intramolecular junction is shown in the inset of Fig. 2I. Given that the Fermi level of a semiconducting nanotube is equal to that of a metallic nanotube (*19*), there is no band bending when the bias is zero. When a positive potential is applied to the left terminal, the Fermi levels of the semiconducting channel and the right metallic tube are raised. There will be a Schottky barrier at the right junction for the electron transport. Because the junctions are symmetric, there will be a Schottky barrier under forward bias and another one under reverse bias. The *I-V* characteristic with V_G at zero (Fig. 2I) is symmetric for the negative and positive biases, with the

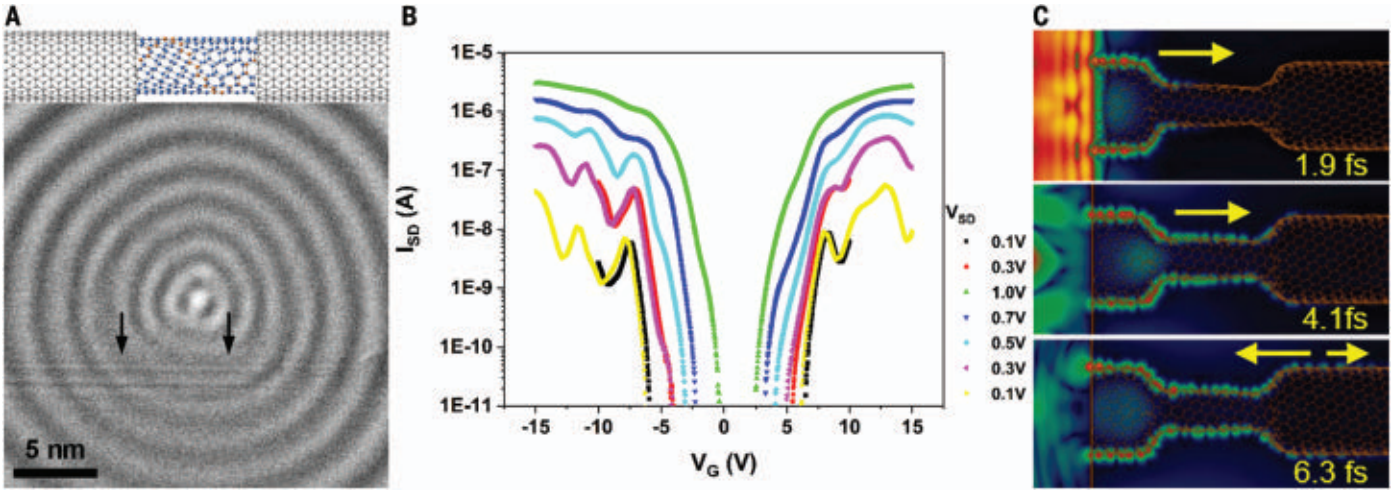


Fig. 3. Room-temperature quantum interference. (**A**) Schematic and TEM image of a SWCNT junction with the kinks marked by arrows. (**B**) I_{SD}-V_G transport curves, showing oscillation conductance peaks at ON state. (**C**) Atomic model of a (22,4)/(9,4)/(22,4) junction with snapshots of the time evolution of an electron wave packet.

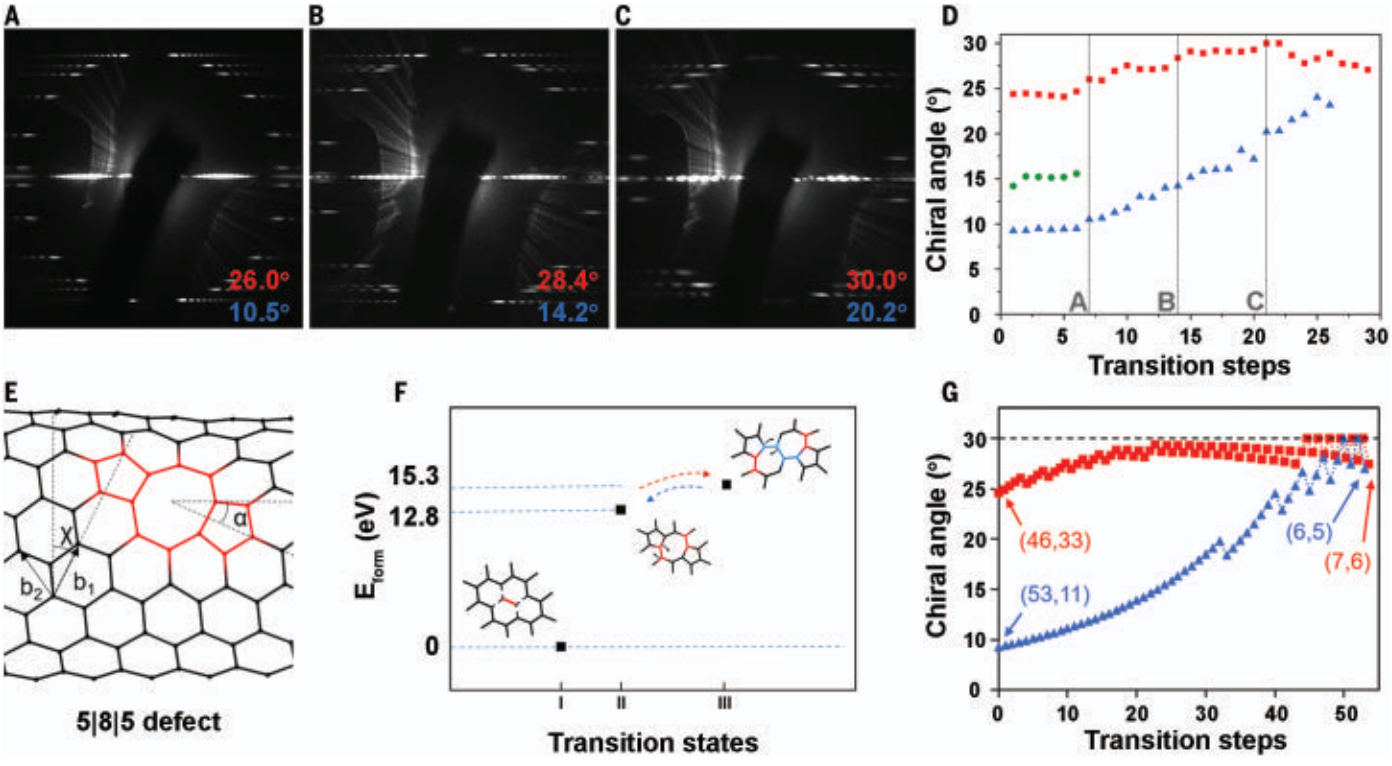


Fig. 4. Chirality transition dynamics. (**A** to **C**) NBED patterns of a CNT during chirality transitions, with the chiral angles marked. (**D**) Changes of the chiral angles revealing a converging trend toward large angles. (**E**) Atomic structure of a (10,7) SWCNT with the basic vectors $\mathbf{b_1}$ and $\mathbf{b_2}$, the chiral angle χ, and misorientation angle α for a 5|8|5 defect. (**F**) Schematics and formation energies (E_{form}) of a pair of dislocations, including the sublimation of a carbon dimer and bond rotation steps. (**G**) Predicted changes of nanotube chiral angles during chirality transitions.

current near zero in the lower bias region and linearly increasing in the higher bias region.

The analysis of the transistor performance under a V_{SD} of 0.5 V (table S1) shows that the SS values are in the range of ~0.18 to ~1.7 V/dec, with the shorter-channel transistors showing greater SS than those with longer channels (fig. S6A). The I_{ON} could reach ~1 μA and seemed to be less dependent on the channel length (fig. S6B). The I_{ON}/I_{OFF} ratio reached ~10^5 for the devices with a smaller SS and therefore more-efficient gate coupling (fig. S6C). While the transistors fabricated in this work showed better performance than reported suspended transistors, the properties are not as good as those of state-of-the-art nanotube transistors fabricated on a substrate. The main disadvantage is the much larger SS, owing to the lower dielectric constant of the vacuum gap in our devices compared with the thin high-κ dielectric film in the supported

transistors, for which SS values near the theoretical limit of 0.06 V/dec have been reported (*3*, *4*, *23*–*25*). Although the suspended nanotube transistors fabricated inside the TEM are ideal for in situ manipulation, characterization, and measurements, such configuration is not a realistic way to produce CNT field-effect transistors for large-scale applications.

We next compared the transfer characteristics of the transistors with channel lengths of ~19.8, ~2.8, and ~0.55 nm (Fig. 2G and fig. S7). Although the I_{ON} current was similar (~1 μA) for the three devices with a V_{SD} of 0.5 V, I_{OFF} was <0.01 nA for the 19.8-nm-channel device and ~ 0.2 nA for the 2.8-nm-channel transistor (Fig. 2H). The device with the channel length of 0.55 nm could not be switched off by V_G, an effect that could have been caused by quantum tunneling at such a short distance. We observed resonant conductance in the gap region for the 2.8-nm-channel transistor, indicating the transition from a semiconductor with continuous energy bands to molecules with discrete energy states. Thus, as the size of the nanotube transistors was reduced, quantum effect and tunneling behavior became more pronounced.

The nanometer length of the chirality-transformed CNT channel creates a quantum mechanically confined region in the axial direction in addition to the circumferential direction. In Fig. 3A, a TEM image and schematic show a SWCNT segment with the length of ~8.1 nm, with small kinks marked by arrows. Oscillations of conductance at the ON state were observed in the V_G-dependent measurements (Fig. 3B). The oscillation peaks gradually smeared out as V_{SD} was increased from 0.1 to 1.0 V, an effect likely caused by a screening effect of the larger I_{SD} and charge density. When V_{SD} was again decreased, the oscillation peaks recovered at the same positions. Such a phenomenon can be understood as Fabry-Pérot interference (*26*, *27*).

The diameter, *d*, of the narrow section was 0.86 nm, which would result in a bandgap of 0.81 eV (E_G = 0.7 eV/*d*). With a V_G gap of 12.5 V, the gate efficiency is 0.065 $\left(\alpha = \frac{E_G}{V_G}\right)$. With a period of conductance V_P of 3.8 V, the energy spacing is calculated as 0.25 eV ($\Delta E = \alpha V_P$). The length of the oscillator is calculated to be 9.1 nm ($L = \frac{hv_F}{2\Delta E}$, where h is Planck's constant and v_F is the Fermi velocity), near the measured length (~8.1 nm). We calculated the propagation process of an electron wave in a nanotube junction using wave packet (WP) dynamics. Snapshots of WP propagation through a (22,4)/(9,4)/(22,4) junction (Fig. 3C and movie S1) revealed the transmittance, reflection, and interference of the electron waves. Spatial and spectral distributions of the wave functions showed the localized states and energy-dependent channels (fig. S8). The observation of quantum interference at room temperature resulted from the large energy spacing from the short chirality-altered segment and from the reduced electron scattering at the covalently bonded nanotube junctions.

The detailed chiral transition processes in a triple-wall CNT (TWCNT) were analyzed by nanobeam electron diffraction (NBED), with intermediate states shown in Fig. 4, A to C. We analyzed the spacings of the layer lines in the NBED patterns to calculate the chiral angles of each wall at each step (Fig. 4D) and found a trend for the chiral angles to increase to the larger angle region near the armchair type of chirality. For a shell with an initial chiral angle of ~9.2°, the chiral angle gradually increased to ~23.2°. For a shell with an initial chiral angle of ~24.4°, the chiral angle increased to 30° and then fluctuated in the large angle region (~27° to ~30°).

Generally, plastic deformation and chirality transformations of CNTs have been attributed to dislocation activity (*14*). In theory, two mechanisms have been discussed for the nucleation of dislocations in CNTs, that is, bond flip driven by stress and thermally activated carbon dimer evaporation (*15*). The chirality changes due to the gliding of dislocations nucleated from bond flip and associated Stone-Thrower-Wales defects (*28*, *29*) have been studied in detail to predict a decreasing trend for the chiral angles to near-zigzag-type chirality owing to the energetically favorable (0,1) dislocations (*14*, *15*), which is the opposite of the experimentally observed increasing trend for the chiral angles in this work. As mentioned, a near-transition condition was used in our experiments so that the CNTs were exposed to a high temperature from Joule heating and accompanied by a slow elongation in a quasi-static process. Therefore, we considered the orientation-dependent formation energy of the dislocations generated by the thermally activated evaporation of carbon dimers (C_2) and associated formation of 5|8|5 defects (Fig. 4E). The 5|8|5 defects can yield 5|7 dislocation cores responsible for the chirality transitions (Fig. 4F). The calculated formation energy of 5|8|5 defects (fig. S9) indicated that at a lower chiral angle, the (1,0) dislocation was energetically favored (fig. S10). Therefore, nanotubes with initially small chiral angles tend to increase the chiral angle during the transitions. In a higher-angle region, the formation energies of 5|8|5 defects associated with (1,0) and (0,1) dislocations became close, which explains the fluctuation behavior of the chiral angles in the high-angle region. The ratio of probabilities of forming (1,0) and (0,1) dislocations was calculated to predict the chirality dynamics (Fig. 4G), showing a converging trend toward larger angles, consistent with the experimental observations.

REFERENCES AND NOTES

1. R. Saito, M. Fujita, G. Dresselhaus, M. S. Dresselhaus, *Appl. Phys. Lett.* **60**, 2204–2206 (1992).
2. A. D. Franklin, Z. Chen, *Nat. Nanotechnol.* **5**, 858–862 (2010).
3. C. Qiu *et al*., *Science* **355**, 271–276 (2017).
4. Q. Cao, J. Tersoff, D. B. Farmer, Y. Zhu, S.-J. Han, *Science* **356**, 1369–1372 (2017).
5. M. M. Shulaker *et al*., *Nature* **501**, 526–530 (2013).
6. G. Hills *et al*., *Nature* **572**, 595–602 (2019).
7. M. S. Arnold, A. A. Green, J. F. Hulvat, S. I. Stupp, M. C. Hersam, *Nat. Nanotechnol.* **1**, 60–65 (2006).
8. J. R. Sanchez-Valencia *et al*., *Nature* **512**, 61–64 (2014).
9. F. Yang *et al*., *Nature* **510**, 522–524 (2014).
10. S. Zhang *et al*., *Nature* **543**, 234–238 (2017).
11. L. Chico, V. H. Crespi, L. X. Benedict, S. G. Louie, M. L. Cohen, *Phys. Rev. Lett.* **76**, 971–974 (1996).
12. Z. Yao, H. W. C. Postma, L. Balents, C. Dekker, *Nature* **402**, 273–276 (1999).
13. M. Ouyang, J.-L. Huang, C. L. Cheung, C. M. Lieber, *Science* **291**, 97–100 (2001).
14. B. I. Yakobson, *Appl. Phys. Lett.* **72**, 918–920 (1998).
15. F. Ding, K. Jiao, M. Wu, B. I. Yakobson, *Phys. Rev. Lett.* **98**, 075503 (2007).
16. J. Y. Huang *et al*., *Nature* **439**, 281–281 (2006).
17. D.-M. Tang *et al*., *Ultramicroscopy* **194**, 108–116 (2018).
18. E. S. Penev, V. I. Artyukhov, B. I. Yakobson, *ACS Nano* **8**, 1899–1906 (2014).
19. A. A. Odintsov, *Phys. Rev. Lett.* **85**, 150–153 (2000).
20. M. Muoth *et al*., *Nat. Nanotechnol.* **5**, 589–592 (2010).
21. A. D. Franklin *et al*., *Nano Lett.* **12**, 758–762 (2012).
22. S. Heinze *et al*., *Phys. Rev. Lett.* **89**, 106801 (2002).
23. A. Javey *et al*., *Nat. Mater.* **1**, 241–246 (2002).
24. Y. Lu *et al*., *J. Am. Chem. Soc.* **128**, 3518–3519 (2006).
25. A. Javey *et al*., *Nano Lett.* **4**, 447–450 (2004).
26. W. Liang *et al*., *Nature* **411**, 665–669 (2001).
27. A. Javey, J. Guo, Q. Wang, M. Lundstrom, H. Dai, *Nature* **424**, 654–657 (2003).
28. P. A. Thrower, *Chem. Phys. Carbon* **5**, 217–320 (1969).
29. A. J. Stone, D. J. Wales, *Chem. Phys. Lett.* **128**, 501–503 (1986).

ACKNOWLEDGMENTS

The authors thank P. Thrower (Penn State) for the constructive advice. V.A.D., D.G.K., and P.B.S. are grateful to P. Vancsó (KFKI) for fruitful discussions. The authors are grateful to a supercomputer cluster at NUST "MISIS" provided by the Materials Modeling and Development Laboratory and to the Joint Supercomputer Center of the Russian Academy of Sciences. **Funding:** This work was supported by JSPS KAKENHI (grants JP20K05281, JP25820336, JP18H05329, JP19H02543, JP20H00220, and JP20KK0114); JST, CREST grant JPMJCR20B5, Japan; the Ministry of Science and Technology of China (grant 2016YFA0200101); and the National Natural Science Foundation of China (NSFC) (grants 52188101, 51625203, and 52130209). S.V.E. and P.B.S. acknowledge financial support from the Ministry of Education and Science of the Russian Federation in the framework of the Increase Competitiveness Program of NUST MISIS (K2-2020-023). D.G.K. and V.A.D. acknowledge support from the Ministry of Science and Higher Education of the Russian Federation (project №01201253304). D.G. is grateful to the Australian Research Council (ARC) for granting Laureate Fellowship FL160100089. **Author contributions:** D.-M.T. conceived of and conducted the in situ TEM experiments. S.V.E. and P.B.S. performed theoretical calculations on the chirality dynamics. V.A.D. and D.G.K. performed theoretical calculations on the quantum transport. O.C. made the LabVIEW program. S.J., L.Z., P.-X.H., C.L., and H.-M.C. performed the growth of SWCNTs through floating catalyst CVD. G.C. and D.N.F. performed the growth of SWCNTs through super-growth CVD. Y.Z., R.X., and S.M. performed the growth of SWCNTs through ethanol CVD. X.Z. and F.-C.H. made probes. N.K., M.M., Y.N., F.U., and M.T. assisted with TEM characterizations. C.L., H.-M.C., P.B.S., Y.B. and D.G. oversaw the project. All authors discussed the results and revised the manuscript. **Competing interests:** None declared. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper or the supplementary materials. Requests for carbon nanotubes under a materials transfer agreement should be made to D.-M.T.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abi8884
Materials and Methods
Figs. S1 to S10
Table S1
References (*30*–*46*)

Movie S1

7 April 2021; accepted 2 November 2021
10.1126/science.abi8884

CORONAVIRUS

Molecular basis of immune evasion by the Delta and Kappa SARS-CoV-2 variants

Matthew McCallum[1], Alexandra C. Walls[1], Kaitlin R. Sprouse[1], John E. Bowen[1], Laura E. Rosen[2], Ha V. Dang[1], Anna De Marco[3], Nicholas Franko[4], Sasha W. Tilles[5], Jennifer Logue[4], Marcos C. Miranda[1,6], Margaret Ahlrichs[1,6], Lauren Carter[1,6], Gyorgy Snell[2], Matteo Samuele Pizzuto[3], Helen Y. Chu[4], Wesley C. Van Voorhis[5], Davide Corti[3], David Veesler[1,7]*

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) transmission leads to the emergence of variants, including the B.1.617.2 (Delta) variant of concern that is causing a new wave of infections and has become globally dominant. We show that these variants dampen the in vitro potency of vaccine-elicited serum neutralizing antibodies and provide a structural framework for describing their immune evasion. Mutations in the B.1.617.1 (Kappa) and Delta spike glycoproteins abrogate recognition by several monoclonal antibodies via alteration of key antigenic sites, including remodeling of the Delta amino-terminal domain. The angiotensin-converting enzyme 2 binding affinities of the Kappa and Delta receptor binding domains are comparable to the Wuhan-Hu-1 isolate, whereas B.1.617.2+ (Delta+) exhibits markedly reduced affinity.

The ongoing spread of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of the COVID-19 pandemic, is resulting in the continued emergence of variants. The B.1.351 (Beta, β) variant of concern was originally described in South Africa and remains the isolate associated with the greatest magnitude of immune evasion, as measured by reduced neutralizing antibody (Ab) titers in vitro (*1*–*3*). Conversely, the B.1.1.7 (Alpha, α) variant of concern, which was first detected in the United Kingdom, has a modest impact on neutralizing Ab titers but a marked enhancement in angiotensin-converting enzyme 2 (ACE2) receptor binding affinity and transmissibility, relative to the Wuhan-Hu-1 isolate, which led to worldwide dominance in the early months of 2021 (*2*, *4*).

The SARS-CoV-2 spike (S) glycoprotein is exposed at the surface of the virus and mediates entry into host cells; it is the main target of neutralizing antibodies and the focus of most vaccines (*5*, *6*). The S glycoprotein is subdivided into two functional subunits, designated S_1 and S_2, that interact noncovalently after proteolytic cleavage by furin during synthesis (*5*, *7*, *8*). The S_1 subunit contains the receptor binding domain (RBD), which engages the receptor ACE2 (*5*, *7*, *9*, *10*), and the N-terminal domain (NTD), which recognizes attachment factors (*11*–*13*). The S_2 subunit contains the fusion machinery and undergoes large-scale conformational changes to drive fusion of the virus and host membranes (*14*), enabling genome delivery and initiation of infection. Abs that bind to specific sites on the RBD (*15*–*22*), the NTD (*23*–*26*), or the fusion machinery stem helix (*27*–*31*) interfere with receptor attachment or membrane fusion. Serum neutralizing Ab titers are a correlate of protection against SARS-CoV-2 in nonhuman primates (*32*–*35*).

In late 2020, B.1.617 variants including B.1.617.1 (Kappa, κ) and B.1.617.2 (Delta, δ) were first detected in India and caused devastating epidemics before spreading globally (*36*, *37*). The B.1.617.1 variant S harbors T95I (Thr^{95}→Ile), G142D, E154K, L452R, E484Q, D614G, P681R, and Q1071H substitutions, whereas the B.1.617.2 variant S carries T19R, G142D, E156G, L452R, T478K, D614G, P681R, and D950N substitutions and a deletion of residues 157 and 158 (157-158del) (table S1). Most of these mutations localize to the RBD (residues 328 to 531) and NTD (residues 14 to 305), which are the major targets of neutralizing Abs in convalescent and vaccinated individuals (*18*, *38*), raising concerns about the efficacy of available vaccines and therapeutic monoclonal antibodies (mAbs) against these variants. Moreover, the K417N mutation was detected in the B.1.617.2 lineage, known as the B.1.617.2+ (Delta plus, δ+) variant, which is shared with the B.1.351 (Beta, β) variant of concern and was previously shown to reduce neutralization potency of some mAbs (*2*, *39*).

To evaluate the effect on neutralization of the mutations in the B.1.617.1, B.1.617.2, and B.1.617.2+ S glycoproteins, we compared vaccine-elicited serum neutralizing activity against ancestral G614 S and these three variant S pseudoviruses. We used a vesicular stomatitis virus (VSV) pseudotyping system (*40*) with ACE2-expressing HEK293T as target cells (*41*). We obtained plasma samples from 40 individuals, specifically from 13, 15, and 12 individuals who had received two doses of Pfizer/BioNtech BNT162b2, two doses of Moderna mRNA-1273, or a single dose of Janssen Ad26.COV2.S, respectively (table S2). The median time after final vaccination was 11 days for Pfizer/BioNtech BNT162b2, 20 days for Moderna mRNA-1273, and 102 days for Janssen Ad26.COV2.S. All three vaccines contain D614 in the S glycoprotein. Twelve sera samples from individuals vaccinated with Janssen Ad26.COV2.S were evaluated, but only eight of these individuals are shown, because of a lack of neutralization when assessed against G614 S pseudovirus for the remaining four samples.

Geometric mean neutralizing Ab titers (GMTs) for the Pfizer/BioNtech BNT162b2–elicited plasma were reduced 3.0-, 2.4-, and 4.1-fold for B.1.617.1, B.1.617.2, and B.1.617.2+ S (GMT: 86, 110, and 63), respectively, compared with G614 S (GMT: 260) (Fig. 1A and figs. S1 and S2). Moderna mRNA-1273–elicited plasma GMTs were reduced 4.1-, 2.6-, and 9.5-fold for B.1.617.1, B.1.617.2, and B.1.617.2+ S (GMTs 170, 270, and 74), respectively, compared with G614 S (GMT: 700) (Fig. 1B and figs. S1 and S2). The average neutralization potency of the Janssen Ad26.COV2.S–elicited plasma was reduced 3.0-, 2.4-, and 3.5-fold for B.1.617.1, B.1.617.2, and B.1.617.2+ S (GMT: 10, 13, and 8.8), respectively, compared with G614 S (GMT: 31) (Fig. 1C and figs. S1 and S2). These data demonstrate that all three variants lead to reductions in neutralization potency from vaccine-elicited Abs, with B.1.617.2+ causing the greatest decrease, on par with what was observed for the B.1.351 (Beta) variant of concern (*1*, *3*, *42*). Furthermore, polyclonal Abs from half of the Janssen-vaccinated individuals evaluated completely lost their ability to neutralize one or multiple variants in our assay, likely as a result of the moderate GMTs against G614 S pseudovirus, which concurs with recent in vitro and in vivo data (*43*, *44*).

Although the B.1.617.1 and B.1.617.2+ variants have greater ability to evade vaccine-elicited Ab neutralization than B.1.617.2—a feature that might increase opportunities for breakthrough infections (*37*, *45*, *46*)—the B.1.617.2 variant became dominant worldwide by June 2021 (Fig. 1, D to F). B.1.617.1 and B.1.617.2 sequences were first observed in late 2020, whereas B.1.617.2+ sequences were first detected in April 2021. As of 30 September 2021, 6806 B.1.617.1, 1,415,761 B.1.617.2, and 3966 B.1.617.2+ sequenced genomes were deposited in GISAID (table S1). The high incidence of B.1.617.2 is consistent with recent studies showing that the B.1.617.2 variant has enhanced transmissibility, replication kinetics, and viral loads in oropharyngeal and nose-throat swabs of infected individuals relative

[1]Department of Biochemistry, University of Washington, Seattle, WA 98195, USA. [2]Vir Biotechnology, San Francisco, CA 94158, USA. [3]Humabs Biomed SA, a subsidiary of Vir Biotechnology, 6500 Bellinzona, Switzerland. [4]Division of Allergy and Infectious Diseases, University of Washington, Seattle, WA 98195, USA. [5]Center for Emerging and Re-emerging Infectious Diseases, Division of Allergy and Infectious Diseases, Department of Medicine, University of Washington School of Medicine, Seattle, WA 98195, USA. [6]Institute for Protein Design, University of Washington, Seattle, WA 98195, USA. [7]Howard Hughes Medical Institute, University of Washington, Seattle, WA 98195, USA.
*Corresponding author. Email: dveesler@uw.edu

to the ancestral Wuhan-Hu-1 virus and other variants (*47*–*49*).

We determined cryo–electron microscopy (cryo-EM) structures of the B.1.617.1 and B.1.617.2 S glycoproteins to gain insights into their reduced antibody sensitivity and provide a structural framework to understand the role of shared and distinct mutations. The B.1.617.1 S structure—harboring the HexaPro stabilizing mutations (F817P, A892P, A899P, A942P, K986P, and V987P) (*50*), a native furin cleavage site, and with an engineered disulfide stapling the RBD in the closed conformation (S383C and D985C) (*51*)—was determined in complex with S309 [an RBD-targeted neutralizing mAb (*15*)] and S2L20 [an NTD-targeted non-neutralizing mAb (*23*, *52*)] at 2.4-Å resolution with all three RBDs closed (Fig. 2A; fig. S3, A to C; and table S3). Local classification and refinements of S309 bound to the B.1.617.1 RBD were used to account for conformational dynamics and improve local resolution of this region to 3.3 Å (Fig. 2, A and C, and fig. S3A). The B.1.617.2 S structure—harboring the F817P, A892P, A899P, A942P, V987P, Y707C, and T883C vFLIP prefusion-stabilizing mutations (*53*) and a native furin cleavage site—was obtained in complex with S2M11 [an RBD-targeted neutralizing mAb (*19*)] and S2L20 at 2.4-Å resolution with all three RBDs closed (Fig. 2B; fig. S3, D to F; and table S3).

The RBD is the main target of serum neutralizing activity in convalescent and vaccinated individuals and comprises several antigenic sites recognized by neutralizing Abs with a range of neutralization potencies and breadth (*15*–*18*, *38*, *54*, *55*). E484Q, L452R, and T478K mutations are part of antigenic site I, which we previously showed to be immunodominant (*16*–*18*). The B.1.617.1 S and B.1.617.2 S structures resolve the complete RBD and provide high-resolution blueprints of the residue substitutions found in these two variants (Fig. 2, C and D). In both structures, the L452R side chain, which is part of antigenic site Ib (*18*), is oriented identically (and modeled in the same rotameric configuration) to what we previously observed in the B.1.427/B.1.429 S structure (*52*) (Fig. 2, E and F). The L452R mutation reduces neutralization mediated by some clinical mAbs, such as bamlanivimab (LY-CoV555) and regdanvimab (CT-P59), owing to steric alteration of this antigenic site that is incompatible with binding (Fig. 2, E and F) (*52*). The B.1.617.1 E484Q RBD mutation is located within the receptor binding motif, at the boundary between antigenic sites Ia and Ib (*18*), and could affect neutralization from mAbs recognizing both subsites. The E484Q mutation is conservative but could alter electrostatic interactions, as it substitutes the side-chain carboxylic group with an amide group through replacement of an oxygen with a nitrogen atom. Residue 484 participates in the epitopes recognized by bamlanivimab (*20*, *56*) and casirivimab (REGN10933) (*57*, *58*), and both mAbs have dampened binding or neutralization potency against E484Q-harboring mutants, likely owing to disruption of electrostatic interactions (*56*, *58*). The B.1.617.2 T478K RBD mutation is also located at the boundary between antigenic sites Ia and Ib but does not affect binding to mAbs that have received an emergency use authorization in the US (*56*, *59*, *60*) (Fig. 2D).

Both the B.1.351 (Beta) and P.1 (Gamma) variants of concern harbor the E484K mutation, which is associated with a reduction of

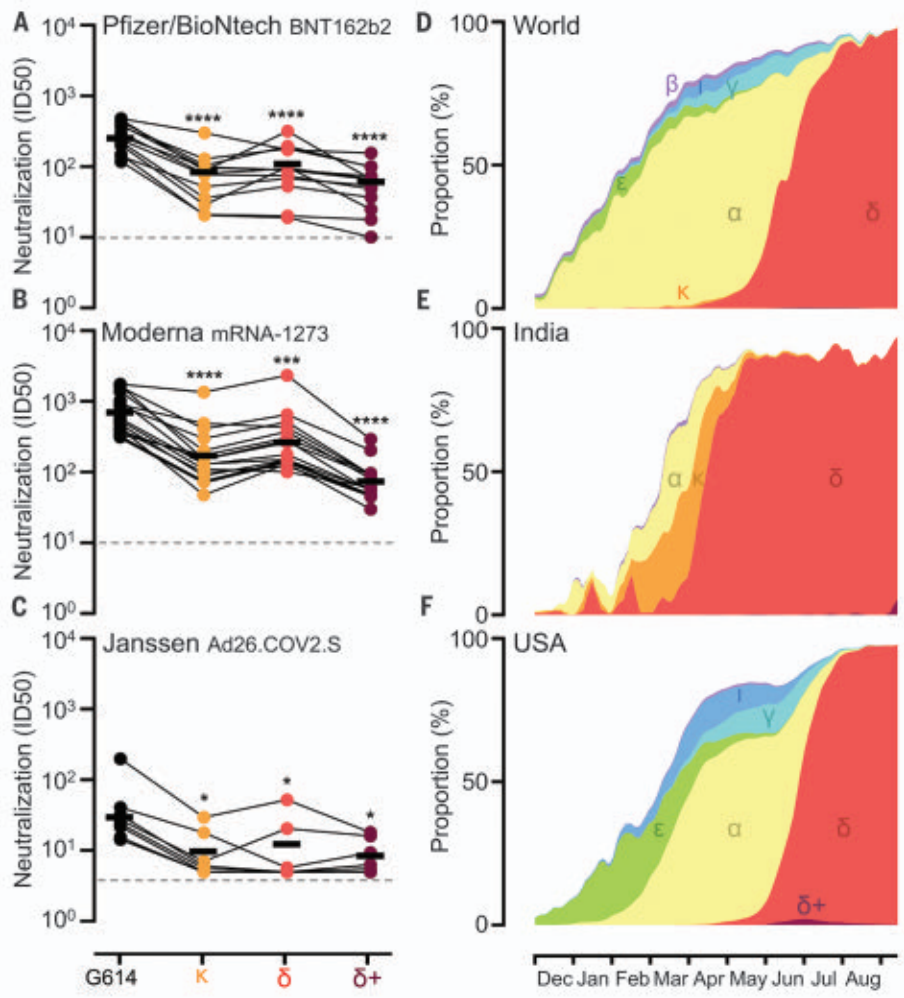


Fig. 1. Neutralization of pseudotyped viruses harboring SARS-CoV-2 G614 S, B.1.617.1 S, B.1.617.2 S, or B.1.617.2+ S and incidence of variants. (**A** to **C**) Pairwise connected half-maximal inhibition dilution (ID50) for each individual against each variant for Pfizer/BioNtech BNT162b2 (A), Moderna mRNA-1273 (B), or Janssen Ad26.COV2.S (C) vaccinee sera (fig. S1 and table S2). Geometric means are shown as thick black horizontal lines. Data are an average of $n = 2$ replicates and are representative of at least two independent assays with distinct batches of pseudoviruses. Dashed line indicates the limit of detection for the assay. Means were compared against G614 by two-way analysis of variance (Dunnett's test); $^{*}P < 0.05$; $^{***}P < 0.001$; $^{****}P < 0.0001$. (**D** to **F**) Incidence (7-day average) of variants of concern and variants of interest as a proportion of viruses sequenced in the world (D), India (E), and the USA (F) deposited to GISAID (analyzed using outbreak.info) from 1 December 2020 to 30 September 2021. B.1.1.7 (Alpha, α), B.1.351 (Beta, β), P.1 (Gamma, γ), B.1.617.2 (Delta, δ) including AY.3-AY.31, B.1.526 (Iota, ι), B.1.427/B.1.429 (Epsilon, ε), B.1.617.1 (Kappa, κ), and B.1.617.2+ (Delta+, δ+) including AY.1 and AY.2 are shown in yellow, light purple, cyan, red, blue, green, orange, and dark purple, respectively.

vaccine-elicited neutralizing Ab titers and mAb neutralization (*39*, *42*, *61*). The E484Q mutation has also been reported to reduce serum neutralizing Ab titers alone or in combination with the L452R mutation (*42*, *58*, *62*). The effect of the T478K mutation on polyclonal Ab responses has not been characterized as thoroughly as the L452R and E484Q substitutions. The recurrent emergence of mutants carrying residue substitutions at positions 452 and 484 underscores the apparent hyperfocusing of neutralizing antibody responses at these antigenic subsites and provides a possible explanation for their acquisition in multiple SARS-CoV-2 isolates. Our structural data integrated with previous studies (*15*, *48*, *60*, *63*) emphasize that S309 is a variant-proof neutralizing mAb that is thus far resilient to the emergence of SARS-CoV-2 variants.

We determined the impact of the RBD mutations present in B.1.617.1, B.1.617.2, and B.1.617.2+ on binding affinity for the ACE2 receptor to assess potential impact on a component of viral transmissibility. The B.1.617.1 (L452R and E484Q) and B.1.617.2 (L452R and T478K) RBDs recognized immobilized ACE2 with roughly similar efficiency to the wild-type RBD, as observed by enzyme-linked immunosorbent assay (ELISA) (Fig. 2G, fig. S4, and table S4). ACE2 binding to the B.1.617.2+ (K417N/L452R/T478K) and B.1.1.7 (N501Y) RBDs was respectively much weaker and much tighter than for any other variants evaluated in our assay, in agreement with the enhanced affinity conferred by N501Y (*4*, *64*) (Fig. 2G and table S4). We confirmed these results using surface plasmon resonance (Fig. 2H, fig. S4, and table S4) and biolayer interferometry (Fig. 2I, fig. S4, and table S4) binding analysis of the monomeric human ACE2 ectodomain to immobilized RBDs, indicating that the B.1.617.1 and B.1.617.2 RBDs interact with ACE2 with roughly comparable affinity to the wild-type RBD whereas the B.1.617.2+ RBD is severely attenuated. These data concur with evaluation of the effect of individual residue mutations on ACE2 binding using deep-mutational scanning of yeast-displayed RBDs (*64*), the positioning of the R452 and K478 side chains away from the ACE2-binding interface, and the conservative nature of the E484Q substitution (Fig. 2, J and K). Finally, these data point to the key contribution of the salt bridge formed between the $K417_{SARS\text{-}CoV\text{-}2}$ and $D30_{ACE2}$ side chains for receptor engagement (Fig. 2, J and K), in agreement with previous studies (*65*, *66*). Because none of the B.1.617.1 and B.1.617.2 RBD mutations increase ACE2 binding markedly and all of them reside in the most immunogenic antigenic site, we suggest that these mutations might have emerged mainly as a result of antibody-mediated selective pressure to reduce immune recognition.

Although multiple antigenic sites are present at the surface of the NTD, a single supersite of vulnerability is targeted by neutralizing Abs elicited upon infection and vaccination (*23*, *24*, *26*). This antigenic supersite (designated site i) comprises the NTD N terminus (residues 14 to 20), a β-hairpin (residues 140 to 158), and a loop (residues 245 to 264). To improve the cryo-EM map resolution of the B.1.617.1 and B.1.617.2 NTDs bound by S2L20 and visualize the structural changes associated with their respective constellation of mutations, we used focused three-dimensional classification and local refinement. Two of the three B.1.617.1 NTD mutations, G142D and E154K, map to the supersite β-hairpin (Fig. 3A and

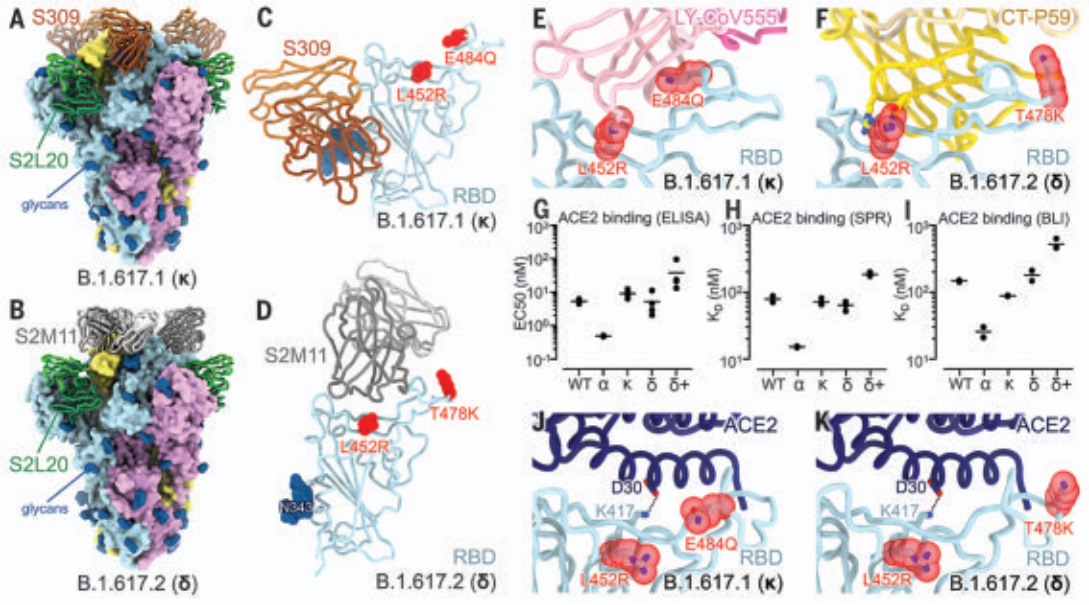


Fig. 2. Cryo-EM structures of the SARS-CoV-2 B.1.617.1 and B.1.617.2 S ectodomain trimers and analysis of ACE2 binding. (**A** and **B**) Structure of the B.1.617.1 (A) and B.1.617.2 (B) S trimer (surface rendering) bound to the S2L20 and S309 (A) or S2M11 (B) Fabs (ribbons). SARS-CoV-2 S protomers are colored pink, cyan, and gold, whereas the S2L20 Fab heavy and light chains are colored dark and light green, respectively. The S309 Fab heavy and light chains are colored dark and light orange, respectively (A). The S2M11 Fab heavy and light chains are colored dark and light gray, respectively (B). Only the Fab variable domains are resolved and therefore modeled in the map. N-linked glycans are rendered as dark blue spheres. 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. (**C**) Zoomed-in view of the S309-bound B.1.617.1 RBD with L452R and E484Q shown as red spheres. (**D**) Zoomed-in view of the S2M11-bound B.1.617.2 RBD with L452R and T478K shown as red spheres. (**E**) Superimposition of the LY-CoV555–bound SARS-CoV-2 RBD structure [Protein Data Bank (PDB) ID 7KMG] on the SARS-CoV-2 B.1.617.1 S cryo-EM structure shows that L452R would clash with the mAb and E484Q would disrupt electrostatic interactions. (**F**) Superimposition of the CT-P59–bound SARS-CoV-2 RBD structure (PDB ID 7CM4) on the SARS-CoV-2 B.1.617.2 S cryo-EM structure shows that L452R would sterically clash with the mAb. (**G**) ELISA binding analysis of the SARS-CoV-2 wild-type, B.1.1.7 (α), B.1.617.1 (κ), B.1.617.2 (δ), and B.1.617.2+ (δ+) RBDs to immobilized human ACE2 ectodomain (residues 1 to 615) shown as half-maximal effective concentrations (EC_{50}). Data from two biological replicates are shown with two to four technical replicates each. (**H**) Surface plasmon resonance (SPR) binding affinity analysis of the human ACE2 ectodomain (residues 1 to 615) for immobilized biotinylated wild-type, B.1.1.7 (α), B.1.617.1 (κ), B.1.617.2 (δ), and B.1.617.2+ (δ+) RBDs. Data from two biological replicates are shown with two to six technical replicates each. (**I**) Biolayer interferometry (BLI) binding analysis of the human ACE2 ectodomain (residues 1 to 615) to immobilized biotinylated SARS-CoV-2 wild-type, B.1.1.7 (α), B.1.617.1 (κ), B.1.617.2 (δ), and B.1.617.2+ (δ+) RBDs. Data from two biological replicates are shown with 1 or 2 technical replicates each. (**J** and **K**) Superimposition of the ACE2-bound SARS-CoV-2 RBD structure (PDB ID 6VW1) on the SARS-CoV-2 B.1.617.1 (J) and B.1.617.2 (K) S cryo-EM structures show that L452R and T478K point away from the interface with ACE2, while K417 contacts D30 from ACE2.

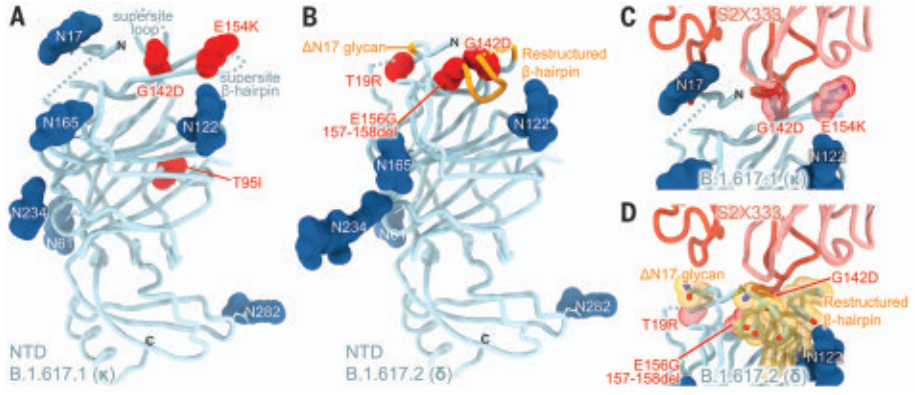


Fig. 3. Remodeling of the NTD antigenic supersite in the B.1.617.1 and B.1.617.2 S variants. (**A** and **B**) Ribbon diagrams of the B.1.617.1 (A) and B.1.617.2 (B) NTDs in the same orientation. Mutated residues are rendered as red spheres and N-linked glycans are shown as dark blue surfaces. Segments with notable structural changes as a consequence of these mutated residues are shown in orange and labeled. (**C** and **D**) Zoomed-in views of the B.1.617.1 (C) and B.1.617.2 (D) NTD antigenic supersites highlighting incompatibility with recognition by the S2X333 mAb (*23*) (used here as an example of prototypical NTD neutralizing mAb). N and C termini are labeled.

table S1), and we have previously shown that G142D abrogates binding of three out of five NTD-specific neutralizing mAbs tested (*23*). The B.1.617.1 T95I substitution occurs outside the antigenic supersite and is unlikely to contribute substantially to immune evasion (Fig. 3A). All of the B.1.617.2 mutations are found within the antigenic supersite: T19R, G142D, E156G, and 157-158del (Fig. 3B). The T19R substitution abrogates the glycosylation sequon at position N17, as supported by the lack of a resolved glycan at this position in the cryo-EM map. We previously showed that T19A, which also removes the N17 glycosylation sequon, decreased binding to four out of five NTD neutralizing mAbs tested (*23*). Residues 156 to 158 participate in the supersite β-hairpin and their mutation or deletion in the B.1.617.2 NTD leads to notable structural remodeling: residues 151 to 159 adopt an α-helical conformation, whereas this segment is mostly β-stranded in the absence of this mutation or deletion (Fig. 3B and fig. S5). On the basis of these findings, we hypothesized that B.1.617.1 and B.1.617.2 variant S glycoproteins would escape recognition by most neutralizing NTD Abs (Fig. 3, C and D).

We therefore evaluated binding of a panel of neutralizing NTD antibodies to variant S ectodomain trimers by ELISA (Fig. 4A and fig. S6). Of the 11 neutralizing mAbs tested, we observed a 10-fold or greater reduction in binding for 8, 10, 10, 10, 3, and 11 mAbs to B.1.617.1, B.1.617.2, B.1.1.7, B.1.351, P.1, and B.1.427/B.1.429 S, respectively, as a result of the introduced mutations and deletions whereas the non-neutralizing S2L20 mAb efficiently recognized all S trimers through binding to antigenic site iv (*23*), confirming proper folding (Fig. 4A). Collectively, these data indicate that NTD-specific neutralizing Abs might exert a selective pressure participating in evolution of the antigenic supersite leading to neutralization escape in numerous variants (*67*–*69*), including B.1.617.1 and B.1.617.2. The diversity of molecular solutions observed in SARS-CoV-2 variants to evade NTD-targeted neutralizing Ab responses further underscores the plasticity of this domain.

S2X303 stood out because of its greater cross-reactivity with variants, including B.1.351, P.1, B.1.617.1, and to a lesser extent B.1.617.2, compared with all other neutralizing mAbs evaluated (Fig. 4A). To provide a structural framework for understanding the S2X303 binding breadth, we characterized its Fab fragment bound to B.1.617.1 S using cryo-EM (Fig. 4B). Focused classification and local refinement of the S2X303-bound NTD yielded a map at 3.5-Å resolution revealing the recognition mode of this mAb (Fig. 4C; fig. S3, G to I; and table S3).

S2X303 recognizes the NTD with an angle of approach almost orthogonal relative to several previously described NTD-specific neutralizing mAbs, and its epitope only partially overlaps with the NTD antigenic supersite (*23*). Specifically, the S2X303 complementary determining region 3–dominated paratope exclusively contacts residues 123 to 125 (along with the glycan at position N122), which are part of the NTD galectin-like distal β sheet, and residues 144 to 154 within the supersite β-hairpin (Fig. 4, D and E). As a result, S2X303 defines a new class of NTD-specific mAbs that differ from S2X333 (*23*), a canonical ultrapotent NTD-specific mAb, and from P008_056 (*70*), which competes with biliverdin binding, with respect to their footprints and angles of approach (Fig. 4E). Moreover, S2X303 exhibits better cross-reactivity compared with all other mAbs and can bind to a large number of variants, although not all of the ones tested (Fig. 4A). Furthermore, we found that although S2X303 neutralized G614 S pseudovirus, it could not neutralize the B.1.617.1 S and B.1.617.2 S pseudoviruses, putatively reflecting enhanced off-rate and reduced binding affinity not detected by ELISA (fig. S7).

Mutations found in the B.1.617.1 and B.1.617.2 variants mediate immune evasion by eroding infection- and vaccine-elicited serum neutralizing Ab titers due to structural alteration present in major antigenic sites within the RBD and NTD (*37*, *71*, *72*). Neither of the fusion machinery B.1.617.1 Q1071H or the B.1.617.2 D950N substitutions are part of epitopes known to be recognized by neutralizing Abs. The B.1.617.2+ variant is associated with a severe dampening of neutralizing Ab titers owing to the additional K417N RBD substitution, also found in the B.1.351 variant of concern (*1*, *2*, *39*). However, the deleterious effect of this mutation on ACE2 binding and absence of the compensatory N501Y mutation found in B.1.351 (*4*) might be associated with a fitness cost, putatively explaining the small number of genomes detected for this variant. On the basis of the roughly comparable ACE2 binding affinities of the B.1.617.1, B.1.617.2, and the Wuhan-Hu-1 RBDs, we propose that other factors contribute to the enhanced transmissibility of the B.1.617.2 variant. The P681R mutation found in the B.1.617.1 and B.1.617.2 variants was recently shown to enhance cleavage at the S_1-S_2 boundary, cell-cell fusion, and replication kinetics (*48*, *49*, *72*, *73*). Given that the S_1-S_2 cleavage site is key for transmission and pathogenicity (*74*, *75*), it appears likely that this mutation contributes to the success of these lineages. Furthermore, the L452R mutation, which augments stability and expression (*64*), increases viral replication kinetics relative to the Wuhan-Hu-1 virus (*76*). The enhanced ability of B.1.1.7 to antagonize host innate immunity through up-regulation of orf6 and orf9b was suggested to have contributed to the success of this variant (*77*). It is possible that B.1.617.2 evolved a similar strategy to reach global domination warranting further studies to uncover the contribution of innate immune antagonism to the continued emergence of variants with greater transmissibility.

S309—the parent mAb of sotrovimab, which has received an emergency use authorization from the US Food and Drug Administration—is unaffected by antigenic drift observed in variants of concern and interest, including B.1.617.1 and B.1.617.2(+), owing to recognition of a conserved RBD epitope (*15*, *48*, *60*, *63*), as illustrated here. The recent discovery of multiple additional conserved antigenic sites recognized by RBD-specific mAbs with (near) pan-sarbecovirus neutralizing activity (*16*, *17*, *21*, *78*–*80*) and the anticipated continued emergence of SARS-CoV-2 variants motivate the clinical development and deployment of some of these

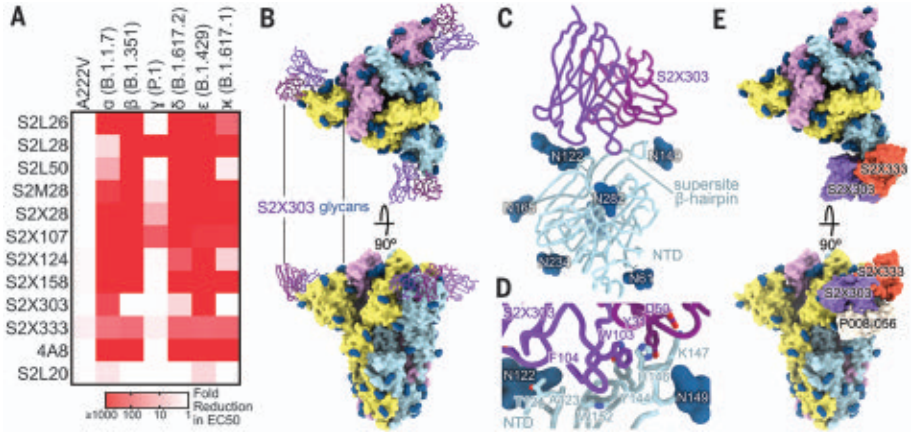


Fig. 4. S2X303 defines a subclass of site i NTD mAbs cross-reacting with several variants. (**A**) Binding of a panel of 11 neutralizing (antigenic site i) and 1 non-neutralizing (S2L20, antigenic site iv) NTD-specific mAbs to recombinant SARS-CoV-2 S variants analyzed by ELISA displayed as a heatmap (relative to wild-type Wuhan-Hu-1 binding). (**B**) Structure of the B.1.617.1 S trimer (surface rendering) bound to the S2X303 Fab fragment (ribbons) shown in two orthogonal orientations. SARS-CoV-2 S protomers are colored pink, cyan, and gold, whereas the S2X303 Fab heavy and light chains are colored dark and light purple, respectively. Only the Fab variable domains are resolved and therefore modeled in the map. N-linked glycans are rendered as dark blue spheres. (**C**) Ribbon diagram of the S2X303-bound SARS-CoV-2 B.1.617.1 NTD. (**D**) Zoomed-in view of the S2X303-bound B.1.617.1 NTD with key residues involved in the interface shown as sticks. (**E**) Structure of the S trimer bound to the S2X303 overlaid with S2X333 and P008-056 antibodies (PDB IDs 7LXW and 7NTC, respectively) shown as a surface rendering. S is colored as in (B); S2X303, S2X333, and P008-056 are shown in purple, orange, and light gray, respectively.

mAbs for prophylaxis and for treatments of unvaccinated individuals or breakthrough infections in at-risk patients. Moreover, next-generation vaccine candidates have recently been described to elicit broad sarbecovirus immunity (*32*, *42*, *81*–*83*), holding the promise to be resilient to the emergence of SARS-CoV-2 variants and of new zoonotic sarbecoviruses. Looking forward, the discovery of broadly neutralizing mAbs targeting the fusion machinery makes tangible the development of a universal β-coronavirus vaccine (*27*–*31*).

REFERENCES AND NOTES

1. S. Cele *et al.*, *Nature* **593**, 142–146 (2021).
2. H. Tegally *et al.*, *Nature* **592**, 438–443 (2021).
3. C. K. Wibmer *et al.*, *Nat. Med.* **27**, 622–625 (2021).
4. D. A. Collier *et al.*, *Nature* **593**, 136–141 (2021).
5. A. C. Walls *et al.*, *Cell* **181**, 281–292.e6 (2020).
6. D. Wrapp *et al.*, *Science* **367**, 1260–1263 (2020).
7. M. Hoffmann *et al.*, *Cell* **181**, 271–280.e8 (2020).
8. M. Hoffmann, H. Kleine-Weber, S. Pöhlmann, *Mol. Cell* **78**, 779–784.e5 (2020).
9. M. Letko, A. Marzi, V. Munster, *Nat. Microbiol.* **5**, 562–569 (2020).
10. P. Zhou *et al.*, *Nature* **579**, 270–273 (2020).
11. S. Wang *et al.*, *Cell Res.* **31**, 126–140 (2021).
12. W. T. Soh *et al.*, *bioRxiv* 2020.11.05.369264 [Preprint] (2020). https://doi.org/10.1101/2020.11.05.369264.
13. F. A. Lempp *et al.*, *Nature* **598**, 342–347 (2021).
14. A. C. Walls *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 11157–11162 (2017).
15. D. Pinto *et al.*, *Nature* **583**, 290–295 (2020).
16. M. A. Tortorici *et al.*, *Nature* **597**, 103–108 (2021).
17. T. N. Starr *et al.*, *Nature* **597**, 97–102 (2021).
18. L. Piccoli *et al.*, *Cell* **183**, 1024–1042.e21 (2020).
19. M. A. Tortorici *et al.*, *Science* **370**, 950–957 (2020).
20. B. E. Jones *et al.*, *Sci. Transl. Med.* **13**, eabf1906 (2021).
21. C. A. Jette *et al.*, *Cell Rep.* **36**, 109760 (2021).
22. C. O. Barnes *et al.*, *Nature* **588**, 682–687 (2020).
23. M. McCallum *et al.*, *Cell* **184**, 2332–2347.e16 (2021).
24. G. Cerutti *et al.*, *Cell Host Microbe* **29**, 819–833.e7 (2021).
25. X. Chi *et al.*, *Science* **369**, 650–655 (2020).
26. N. Suryadevara *et al.*, *Cell* **184**, 2316–2331.e15 (2021).
27. M. M. Sauer *et al.*, *Nat. Struct. Mol. Biol.* **28**, 478–486 (2021).
28. C. Wang *et al.*, *Nat. Commun.* **12**, 1715 (2021).
29. P. Zhou *et al.*, bioRxiv 2021.03.30.437769 [Preprint] (2021). https://doi.org/10.1101/2021.03.30.437769.
30. D. Pinto *et al.*, *Science* **373**, 1109–1116 (2021).
31. G. Song *et al.*, *Nat. Commun.* **12**, 2938 (2021).
32. P. S. Arunachalam *et al.*, *Nature* **594**, 253–258 (2021).
33. K. McMahan *et al.*, *Nature* **590**, 630–634 (2021).
34. D. S. Khoury *et al.*, *Nat. Med.* **27**, 1205–1211 (2021).
35. K. S. Corbett *et al.*, *Science* **373**, eabj0299 (2021).
36. S. Cherian *et al.*, *Microorganisms* **9**, 1542 (2021).
37. P. Mlcochova *et al.*, bioRxiv 2021.05.08.443253 [Preprint] (2021). https://doi.org/10.1101/2021.05.08.443253.
38. A. J. Greaney *et al.*, *Sci. Transl. Med.* **13**, eabi9915 (2021).
39. Z. Wang *et al.*, *Nature* **592**, 616–622 (2021).
40. Y. Kaname *et al.*, *J. Virol.* **84**, 3210–3219 (2010).
41. K. H. D. Crawford *et al.*, *Viruses* **12**, 513 (2020).
42. A. C. Walls *et al.*, *Cell* **184**, 5432–5447.e16 (2021).
43. R. Ukey *et al.*, medRxiv 2021.09.17.21263528 [Preprint] (2021). https://doi.org/10.1101/2021.09.17.21263528.
44. A. Sharma, G. Oda, M. Holodniy, medRxiv 2021.09.23.21263864 [Preprint] (2021). https://doi.org/10.1101/2021.09.23.21263864.
45. T. Kustin *et al.*, *Nat. Med.* **27**, 1379–1384 (2021).
46. P. Y. Chia *et al.*, medRxiv 2021.07.28.21261295 [Preprint] (2021). https://doi.org/10.1101/2021.07.28.21261295.
47. B. Li *et al.*, medRxiv 2021.07.07.21260122 [Preprint] (2021). https://doi.org/10.1101/2021.07.07.21260122.
48. P. Mlcochova *et al.*, *Nature* **599**, 114–119 (2021).
49. Y. Liu *et al.*, bioRxiv 2021.08.12.456173 [Preprint] (2021). https://doi.org/10.1101/2021.08.12.456173.
50. C. L. Hsieh *et al.*, *Science* **369**, 1501–1505 (2020).
51. M. McCallum, A. C. Walls, J. E. Bowen, D. Corti, D. Veesler, *Nat. Struct. Mol. Biol.* **27**, 942–949 (2020).
52. M. McCallum *et al.*, *Science* **373**, 648–654 (2021).
53. E. Olmedillas *et al.*, *bioRxiv* 2021.05.06.441046 [Preprint] (2021). https://doi.org/10.1101/2021.05.06.441046.
54. C. O. Barnes *et al.*, *Cell* **182**, 828–842.e16 (2020).
55. S. J. Zost *et al.*, *Nature* **584**, 443–449 (2020).
56. T. N. Starr, A. J. Greaney, A. S. Dingens, J. D. Bloom, *Cell Rep. Med.* **2**, 100255 (2021).
57. J. Hansen *et al.*, *Science* **369**, 1010–1014 (2020).
58. T. Tada *et al.*, bioRxiv 2021.05.14.444076 [Preprint] (2021). https://doi.org/10.1101/2021.05.14.444076.
59. T. N. Starr *et al.*, *Science* **371**, 850–854 (2021).
60. A. L. Cathcart *et al.*, bioRxiv 2021.03.09.434607 [Preprint] (2021). https://doi.org/10.1101/2021.03.09.434607.
61. K. Wu *et al.*, *N. Engl. J. Med.* **384**, 1468–1470 (2021).
62. I. A. T. M. Ferreira *et al.*, *J. Infect. Dis.* **224**, 989–994 (2021).
63. F. A. Lempp *et al.*, bioRxiv 2021.04.03.438258 [Preprint] (2021). https://doi.org/10.1101/2021.04.03.438258.
64. T. N. Starr *et al.*, *Cell* **182**, 1295–1310.e20 (2020).
65. E. C. Thomson *et al.*, *Cell* **184**, 1171–1187.e20 (2021).
66. M. Yuan *et al.*, *Science* **373**, 818–823 (2021).
67. K. R. McCarthy *et al.*, *Science* **371**, 1139–1142 (2021).
68. B. Choi *et al.*, *N. Engl. J. Med.* **383**, 2291–2293 (2020).
69. V. A. Avanzato *et al.*, *Cell* **183**, 1901–1912.e9 (2020).
70. A. Rosa *et al.*, *Sci. Adv.* **7**, eabg7607 (2021).
71. V.-V. Edara *et al.*, *N. Engl. J. Med.* **385**, 664–666 (2021).
72. D. Planas *et al.*, *Nature* **596**, 276–280 (2021).
73. A. Saito *et al.*, bioRxiv 2021.06.17.448820 [Preprint] (2021). https://doi.org/10.1101/2021.06.17.448820.
74. B. A. Johnson *et al.*, *Nature* **591**, 293–299 (2021).
75. Y. J. Hou *et al.*, *Cell* **182**, 429–446.e14 (2020).
76. C. Motozono *et al.*, *Cell Host Microbe* **29**, 1124–1136.e11 (2021).
77. L. G. Thorne *et al.*, bioRxiv 2021.06.06.446826 [Preprint] (2021). https://doi.org/10.1101/2021.06.06.446826.
78. A. Z. Wec *et al.*, *Science* **369**, 731–736 (2020).
79. C. G. Rappazzo *et al.*, *Science* **371**, 823–829 (2021).
80. D. R. Martinez *et al.*, bioRxiv 2021.04.27.441655 [Preprint] (2021). https://doi.org/10.1101/2021.04.27.441655.
81. D. R. Martinez *et al.*, *Science* **373**, 991–998 (2021).
82. A. A. Cohen *et al.*, *Science* **371**, 735–741 (2021).
83. A. C. Walls *et al.*, *Cell* **183**, 1367–1382.e17 (2020).

ACKNOWLEDGMENTS

We thank H. Tani (University of Toyama) for providing the reagents necessary for preparing VSV pseudotyped viruses. We thank J. L. Mullen, G. Tsueng, A. Abdel Latif, M. Alkuzweny, M. Cano, E. Haag, J. Zhou, M. Zeller, E. Hufbauer, N. Matteson, K. G. Andersen, C. Wu, A. I. Su, K. Gangavarapu, L. D. Hughes, and the Center for Viral Systems Biology outbreak.info (https://outbreak.info/). **Funding:** This study was supported by the National Institute of Allergy and Infectious Diseases (DP1AI158186 and HHSN272201700059C to D.V.), a Pew Biomedical Scholars Award (D.V.), an Investigators in the Pathogenesis of Infectious Disease Award from the Burroughs Wellcome Fund (D.V.), Fast Grants (D.V.), the Bill and Melinda Gates Foundation (OPP1156262 to D.V.), the University of Washington Arnold and Mabel Beckman cryo-EM center and the National Institutes of Health grant S10OD032290 (to D.V.), and grant U01 AI151698 for the United World Antiviral Research Network (UWARN) as part of the Centers for Research in Emerging Infectious Diseases (CREID) Network. D.V. is an investigator of the Howard Hughes Medical Institute. **Author contributions:** M.M. and D.V. conceived of the project. M.M., A.C.W., and D.V. designed the experiments. M.M. analyzed the incidence of variants. M.M. and J.E.B. expressed and purified proteins and performed ELISAs. M.M. carried out cryo-EM sample preparation, data collection, processing model building, and refinement. A.C.W. carried out BLI analysis. L.E.R. performed SPR analysis. A.C.W. and K.R.S. performed pseudovirus neutralization assays and analysis. H.V.D., A.D.M., S.W.T., M.C.M., L.C., M.S.P., H.Y.C., W.C.V.V., and D.C. provided reagents. G.S., D.C., and D.V. supervised the experiments. M.M. and D.V. wrote the manuscript, with input from all authors. **Competing interests:** L.E.R., A.D.M., G.S., M.S.P., and D.C. are employees of Vir Biotechnology, Inc., and may hold shares in Vir Biotechnology, Inc. D.C. is currently listed as an inventor on multiple patent applications, which disclose the subject matter described in this manuscript. A.C.W., G.S., D.C., and D.V. are listed as inventors on patent 49230.03US1 describing the S309 epitope. The Veesler laboratory has received a sponsored research agreement from Vir Biotechnology, Inc. H.Y.C. is a consultant for Merck, Pfizer, Ellume, and the Bill and Melinda Gates Foundation and has received support from Cepheid and Sanofi-Pasteur. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of

interest. **Data and materials availability:** The cryo-EM maps and coordinates have been deposited to the Electron Microscopy Data Bank and Protein Data Bank under the accession numbers listed in table S3. Materials generated in this study will be made available on request after signing a materials transfer agreement with Vir Biotechnology or the University of Washington. 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.org/doi/10.1126/science.abl8506
Materials and Methods
Figs. S1 to S7
Tables S1 to S4
References (*84*–*103*)
MDAR Reproducibility Checklist

11 August 2021; accepted 2 November 2021
Published online 9 November 2021
10.1126/science.abl8506

CORONAVIRUS

Rapid assessment of SARS-CoV-2–evolved variants using virus-like particles

Abdullah M. Syed[1,2]†, Taha Y. Taha[3]†, Takako Tabata[3]†, Irene P. Chen[3,4], Alison Ciling[2], Mir M. Khalid[3], Bharath Sreekumar[3], Pei-Yi Chen[3], Jennifer M. Hayashi[3], Katarzyna M. Soczek[2,5], Melanie Ott[2,3,6]*, Jennifer A. Doudna[1,2,5,7,8,9,10,11]*

Efforts to determine why new severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variants demonstrate improved fitness have been limited to analyzing mutations in the spike (S) protein with the use of S-pseudotyped particles. In this study, we show that SARS-CoV-2 virus-like particles (SC2-VLPs) can package and deliver exogenous transcripts, enabling analysis of mutations within all structural proteins and at multiple steps in the viral life cycle. In SC2-VLPs, four nucleocapsid (N) mutations found universally in more-transmissible variants independently increased messenger RNA delivery and expression ~10-fold, and in a reverse genetics model, the serine-202→arginine (S202R) and arginine-203→methionine (R203M) mutations each produced >50 times as much virus. SC2-VLPs provide a platform for rapid testing of viral variants outside of a biosafety level 3 setting and demonstrate N mutations and particle assembly to be mechanisms that could explain the increased spread of variants, including B.1.617.2 (Delta, which contains the R203M mutation).

The COVID-19 pandemic is a leading cause of death globally, owing to the ongoing emergence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) variants with increased transmissibility. Understanding the molecular determinants of enhanced infectivity is central to vaccine and therapeutic development, but research is hindered because SARS-CoV-2 can be studied only in a biosafety level 3 (BSL-3) laboratory. Furthermore, technical challenges impede efforts to generate mutant infectious clones of SARS-CoV-2 (*1*–*5*). Current studies employ spike (S) protein–pseudotyped lentivirus systems for evaluation of S-mediated ACE2 receptor binding and cell entry (*6*, *7*). However, most mutations in circulating variants occur outside of the S gene and are thus inaccessible by this approach (*8*).

All SARS-CoV-2 variants of interest or concern defined by the World Health Organization contain at least one mutation with >50% penetrance within a seven–amino acid span (199 to 205) in the nucleocapsid (N) protein, which is required for replication, RNA binding, packaging, stabilization, and release (*8*). Despite its functional importance and emergence as a mutational hotspot, the N protein has not been widely studied because of the absence of simple and safe cell-based assays. Biochemical analysis of N has also proven difficult because of its instability and propensity to assemble or phase-separate and to bind RNA nonspecifically (*9*–*11*). To investigate N function, effects of mutations, and other aspects of SARS-CoV-2 biology, we set out to develop a system to package and deliver exogenous RNA transcripts into cells by means of virus-like particles (VLPs).

We reasoned that a process that mimics viral assembly to package and deliver reporter transcripts would simplify the analysis of successful virus production, budding, and entry. Previous studies have shown that coexpression of only the structural proteins of coronaviruses generates VLPs that contain all four structural proteins (*12*–*17*). These VLPs appear to have similar morphology to infectious viruses and have been proposed as vaccine candidates (*18*). A key requirement for such VLPs to deliver reporter transcripts into cells is the recognition of a cis-acting RNA sequence that triggers packaging. During viral assembly, the N protein is thought to recognize one or more RNA structures within open reading frames 1a and 1b (ORF1ab), thus enabling the full viral genome that contains this sequence to be packaged to the exclusion of viral subgenomic and host transcripts (*19*). The identification of such a SARS-CoV-2 cis-acting RNA element is required to create SARS-CoV-2 VLPs (SC2-VLPs) that incorporate and deliver engineered transcripts by this mechanism.

On the basis of the reported packaging sequences for related viruses, including murine hepatitis virus and severe acute respiratory syndrome coronavirus (SARS-CoV), we hypothesized that the SARS-CoV-2 packaging signal might reside within a region termed "T20" (nucleotides 20080 to 22222) that encodes nonstructural proteins 15 and 16 (nsp15 and nsp16) (Fig. 1A) (*16*, *19*–*21*). We designed a transfer plasmid that encodes a luciferase transcript containing T20 within its 3′ untranslated region (UTR). We tested for SC2-VLP production by cotransfecting the transfer plasmid into packaging cells (293T cells) along with plasmids that encode the viral structural proteins (Fig. 1B). Supernatant collected from these cells was filtered and incubated with receiver 293T cells that coexpress SARS-CoV-2 entry factors ACE2 and TMPRSS2 (Fig. 1B). We observed luciferase expression in receiver cells only in the presence of all structural proteins (S, M, N, and E) as well as the T20-containing reporter transcript (Fig. 1C). Substituting any one of the structural proteins or the luciferase-T20 transcript with a luciferase-only transcript decreased luminescence in receiver cells >200- and 63-fold, respectively (Fig. 1C). We also conducted this experiment using Vero E6 cells that endogenously express ACE2 and once again observed robust luciferase expression only in the presence of all five components (fig. S1A).

VLP-mediated transcript delivery required untagged native M protein and a low ratio

[1]Gladstone Institute of Data Science and Biotechnology, San Francisco, CA, USA. [2]Innovative Genomics Institute, University of California, Berkeley, Berkeley, CA, USA. [3]Gladstone Institute of Virology, San Francisco, CA, USA. [4]Biomedical Sciences Graduate Program, University of California, San Francisco, CA, USA. [5]Department of Molecular and Cell Biology, University of California, Berkeley, CA, USA. [6]Department of Medicine, University of California San Francisco, CA, USA. [7]Molecular Biophysics and Integrated Bioimaging Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [8]Howard Hughes Medical Institute, University of California, Berkeley, Berkeley, CA, USA. [9]Department of Chemistry, University of California, Berkeley, Berkeley, CA, USA. [10]California Institute for Quantitative Biosciences (QB3), University of California, Berkeley, Berkeley, CA, USA. [11]Gladstone-UCSF Institute of Genomic Immunology, San Francisco, CA, USA.
***Corresponding author. Email: doudna@berkeley.edu (J.A.D.); melanie.ott@gladstone.ucsf.edu (M.O.)**
†These authors contributed equally to this work.

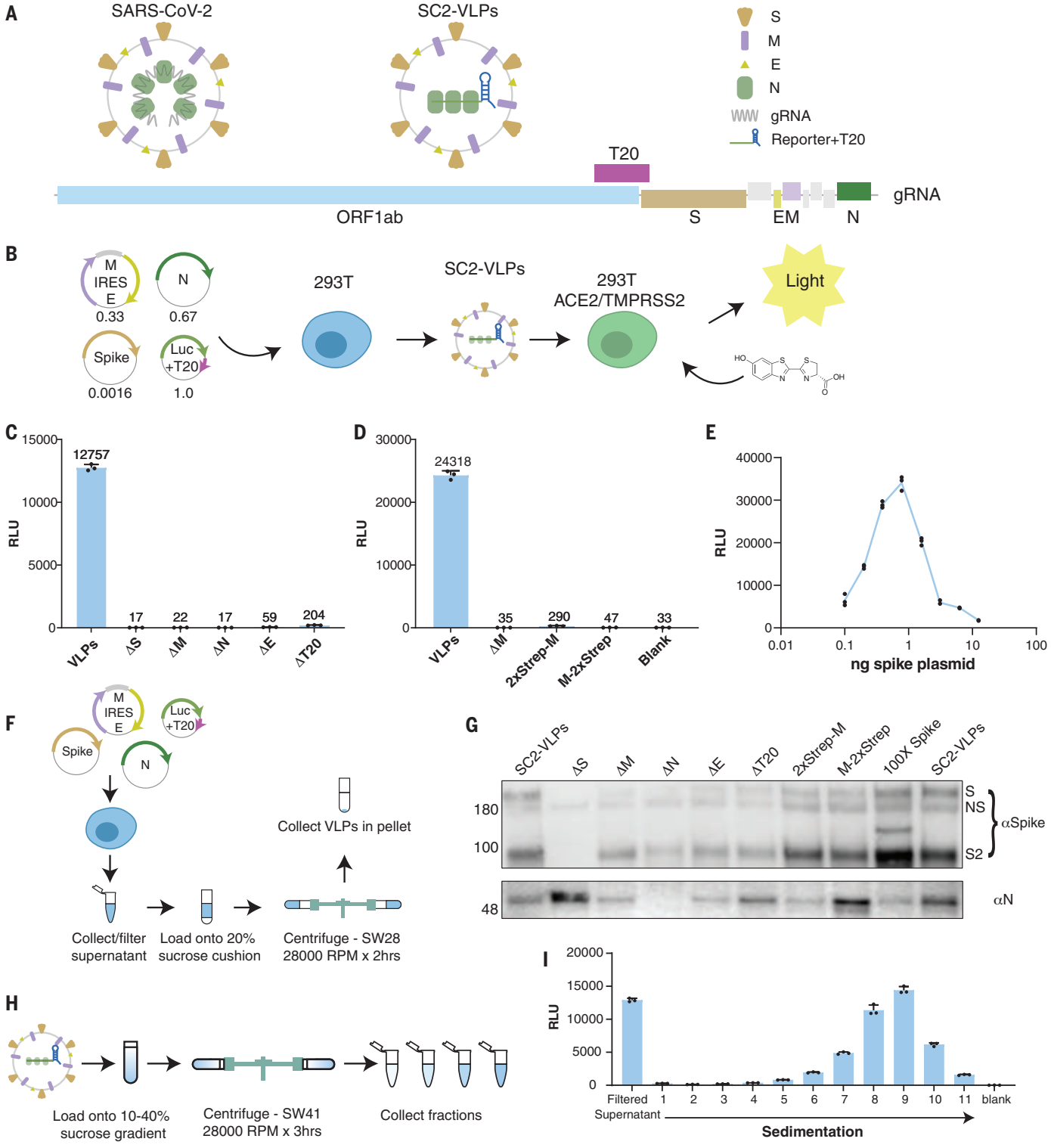


Fig. 1. Design and characterization of SC2-VLPs. (**A**) Schematic of SARS-CoV-2 and SC2-VLP design and location of RNA packaging sequence T20. (**B**) Process for generating and detecting luciferase-encoding SC2-VLPs. Numbers below plasmid maps indicate ratios used for transfection. (**C**) Induced luciferase expression measured in receiver cells (293T-ACE2/TMPRSS2) from "standard" SC2-VLPs containing S, M, N, E, and luciferase-T20 transcript, as well as VLPs lacking each one of the components. (**D**) N- or C-terminal 2× Strep-tag on M abrogates vector-induced luciferase expression. (**E**) Optimal luciferase expression requires a narrow range of spike plasmid concentrations corresponding to $^{1}/_{500}$ of the total plasmid mass used for transfection. (**F**) Schematic for purification of SC2-VLPs. (**G**) Western blot showing S and N in pellets purified from standard SC2-VLPs and conditions that did not induce luciferase expression in receiver cells. (**H**) Schematic for sucrose gradient for separating SC2-VLPs. (**I**) Induced luciferase expression from sucrose gradient fractions of SC2-VLPs. gRNA, genomic RNA; IRES, internal ribosome entry site; Luc, luciferase; RLU, relative light units; S2, cleavage product of S; NS, nonspecific band. Error bars indicate SD with *N* = 3 independent transfections in each case.

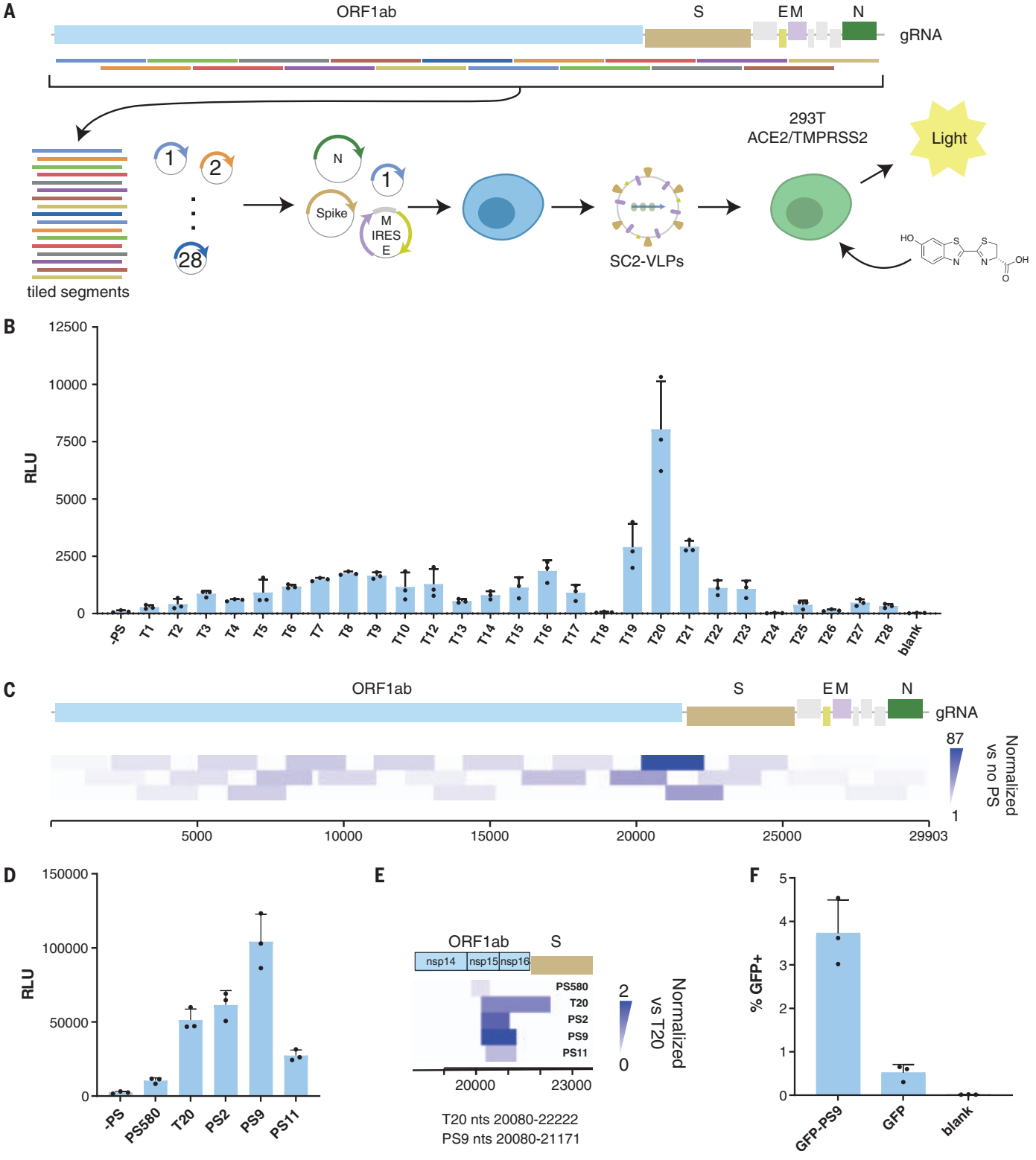


Fig. 2. RNA packaging into SC2-VLPs by SARS-CoV-2 sequences. (**A**) Arrayed screen for determining the location of the optimal sequence for RNA packaging in SC2-VLPs. Two-kilobase tiled segments of the genome were cloned into the 3′ UTR of the luciferase plasmid. (**B**) Induced luciferase expression in receiver cells by SC2-VLPs containing different tiled segments from the SARS-CoV-2 genome. (**C**) Heatmap visualization of the data from (B), showing the locations of tiled segments relative to the SARS-CoV-2 genome. Color intensity indicates luminescence of receiver cells for each tile normalized to expression for a luciferase plasmid containing no insert. (**D**) Smaller segments of the genome were used to locate the optimal RNA packaging sequence. (**E**) Heatmap visualization of the data from (D). (**F**) Flow cytometry analysis of GFP expression in 293T ACE2/TMPRSS2 cells incubated with SC2-VLPs encoding GFP-PS9, GFP (no packaging sequence), or no VLPs. nts, nucleotides. Error bars indicate SD with $N = 3$ independent transfections in each case.

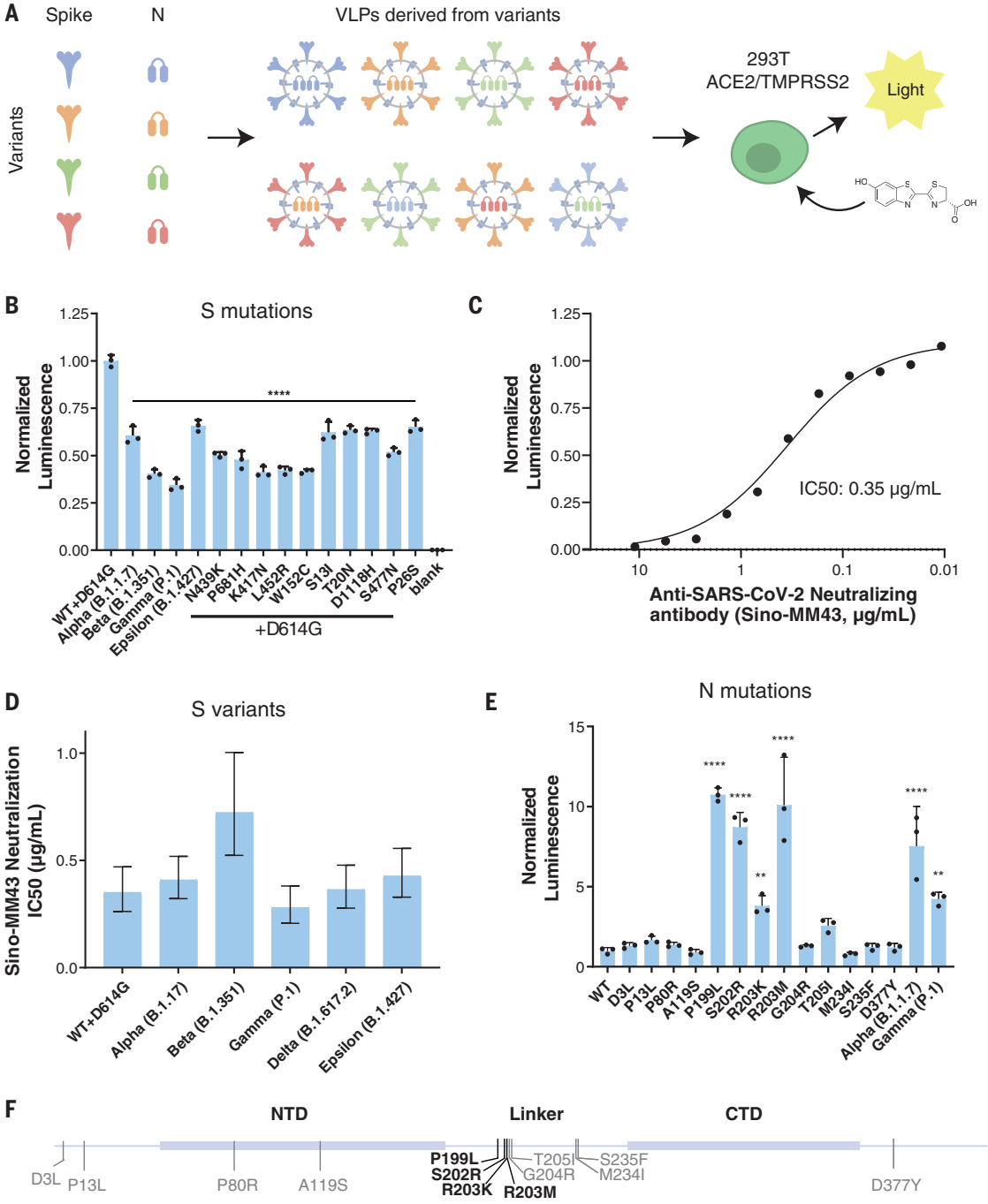


Fig. 3. Effect of mutations in the S and N proteins on SC2-VLP–induced luminescence. (**A**) Schematic for cloning and testing mutations observed in SARS-CoV-2 variants using SC2-VLPs. (**B**) Initial screen of 15 S mutants compared with a reference ancestral S containing the D614G mutation (termed WT). Details of mutations are listed in table S2. (**C**) Neutralization curve for SC2-VLPs generated using ancestral S and neutralized with the anti-S antibody MM43 (SinoBiological, catalog no. 40591). (**D**) Neutralization IC_{50} of S variants using SC2-VLPs and MM43. (**E**) Initial screen of 15 N mutants compared with the reference Wuhan Hu-1 N sequence (WT). Details of mutations are listed in table S3. (**F**) Map of SARS-CoV-2 N domains showing the locations of observed mutations. Mutations that were observed to enhance signal are shown in bold. (B and E) Error bars indicate SD with $N = 3$ independent transfections in each case. Significance was determined by one-way analysis of variance and multiple comparisons using the Holm-Šídák test. $^{**}P < 0.01$; $^{****}P < 0.0001$. (D) Error bars indicate 95% confidence intervals derived from curve fitting in GraphPad Prism. NTD, N-terminal domain; CTD, C-terminal domain. Single-letter abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; R, Arg; S, Ser; T, Thr; W, Trp; and Y, Tyr.

of S expression plasmid relative to the other plasmids (Fig. 1, D and E, and fig. S1B). This is expected because S is expressed at lower levels than other structural proteins during infection (*22*). SC2-VLPs should also require lower S expression compared with pseudovirus systems because S assembles along with the other structural proteins within the endoplasmic reticulum–Golgi intermediate compartment, whereas pseudovirus assays typically require accumulation of S at the plasma membrane followed by random incorporation into budding particles (*23*, *24*). Notably, we detected N and S proteins within pelleted material (Fig. 1, F and G) from multiple conditions that did not yield luciferase expression in receiver cells, suggesting that particles produced under less stringent conditions are not competent for delivering mRNA. These findings suggest that coexpression of structural proteins likely produces VLPs as well as defective particles, and mRNA delivery requires more stringent conditions.

Further analysis showed that SC2-VLPs are stable against ribonuclease A; are resistant to freeze-thaw treatment (fig. S2A); can be concentrated by precipitation, ultrafiltration, and

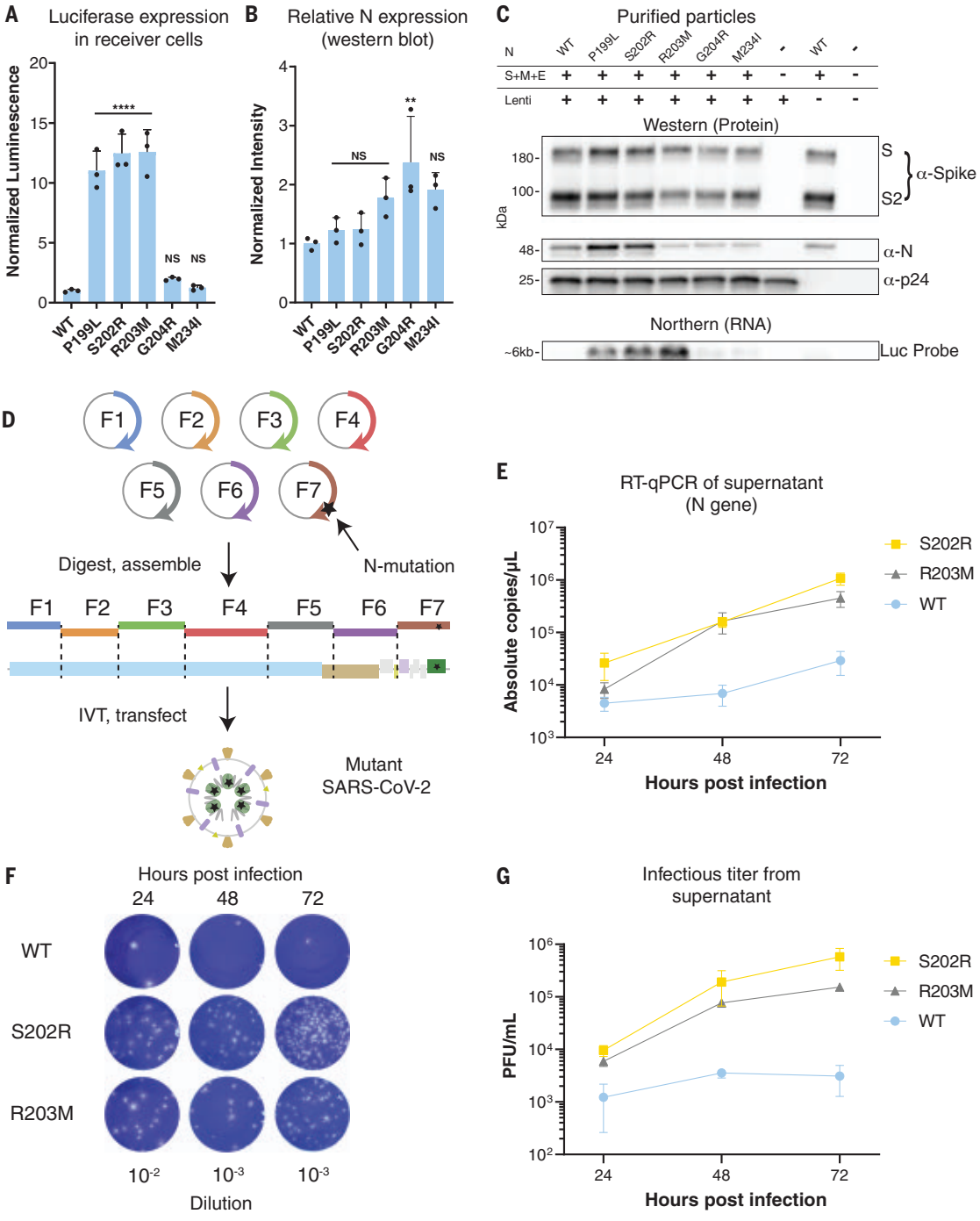


Fig. 4. Impact of mutations in SARS-CoV-2 N on RNA packaging and viral titer. (**A**) Luciferase expression in receiver cells from six N mutants retested after preparation in a larger batch. (**B**) Relative N-expression of selected mutants in packaging cells normalized to WT, with glyceraldehyde-3-phosphate dehydrogenase as a loading control. (**C**) Western blot (protein) and Northern blot (RNA) of VLPs generated using N-mutants purified by ultracentrifugation. Lentivirus was added before ultracentrifugation to allow use of p24 as an internal control. (**D**) Schematic for the reverse genetics system used to generate mutant SARS-CoV-2. (**E**) RT-qPCR of supernatant collected from A549-ACE2 cells infected with WT and mutant SARS-CoV-2 at MOI of 0.1 at 24, 48, and 72 hours after infection. (**F**) Representative plaques and (**G**) quantification of infectious viral titers from the same experiment. Error bars indicate SD with $N = 3$ independent transfections or infections in each case. Significance was determined by one-way analysis of variance and multiple comparisons using Holm-Šídák test. $^{**}P < 0.01$; $^{****}P < 0.0001$; NS, not significant. F1 to F7, fragments 1 to 7; IVT, in vitro transcription; PFU, plaque-forming units.

ultracentrifugation (fig. S2B); and induce transient expression of luciferase (fig. S2C). Analysis of SC2-VLPs by sucrose gradient ultracentrifugation showed that large, dense particles are responsible for inducing luciferase expression (Fig. 1, H and I). These data show that SC2-VLPs are formed under our experimental conditions and deliver selectively packaged transcripts.

Next, we determined the optimal SARS-CoV-2 cis-acting RNA sequence for SC2-VLP–mediated delivery. We generated a library of 28 overlapping tiled segments of 2 kb each (numbered T1 to T28; table S1) from the SARS-CoV-2 genome and inserted them individually into a luciferase-encoding plasmid (Fig. 2A). SC2-VLPs generated using luciferase-encoding plasmids that included any region of ORF1ab produced detectable luminescence, suggesting that SARS-CoV-2 packaging may not rely entirely on one contiguous cis-acting RNA element (Fig. 2, B and C). Furthermore, luciferase-encoding plasmids that included fragments T24 to T28 resulted in lower levels of luciferase expression (Fig. 2, B and C), consistent with the exclusion of subgenomic viral transcripts containing these sequences to minimize production of replication-defective virus particles. Overall, luciferase expression was most efficient using T20 (nucleotides 20080 to 22222) located near the 3′ end of ORF1ab (Fig. 2, B and C) and partially but not completely overlapping with PS580 (nucleotides 19785 to 20348), which was previously predicted to be the packaging signal for SARS-CoV on the basis of structural similarity to known coronavirus packaging signals (*16*).

To further define the minimal cis-acting RNA element sufficient for SC2-VLP–mediated delivery, we tested truncations and additions to T20 as well as PS580 from SARS-CoV. We found that PS580 resulted in lower luciferase expression compared with T20 (Fig. 2, D and E, and fig. S3, A and B). We observed the highest luciferase expression level from SC2-VLPs encoding the nucleotide sequence 20080 to 21171 (termed PS9), and further truncations of this sequence reduced expression (Fig. 2, D and E, and fig. S3, A and B). We used PS9 to generate VLPs encoding green fluorescent protein (GFP) and found that they induced GFP expression in receiver cells (Fig. 2F). These data suggest that PS9 (nucleotides 20080 to 21171) is a cis-acting element that is sufficient for triggering RNA packaging into SC2-VLPs, although it is not currently known whether this sequence is required for the packaging of the SARS-CoV-2 genome.

Compared with pseudoviruses, SC2-VLPs provide a new and more physiological model for testing mutations in all viral structural proteins (S, E, M, and N) for effects on assembly, packaging, and cell entry. Surprisingly, none of the 15 SC2-VLPs generated with S mutant genes—including four with the combined mutations found in the B.1.1.7 (Alpha), B.1.351 (Beta), P.1 (Gamma), and B.1.427 (Epsilon) variants—increased luciferase expression in transduced cells beyond that observed for the original SC2-VLPs (Fig. 3, A and B, and table S2). Because nearly all circulating variants contain the S D614→G (S:D614G) mutation, we compared all mutants to the ancestral S protein modified to include G614 (termed WT+D614G). Minor changes in S expression between mutants may be a confounding factor because SC2-VLPs mediate luciferase expression optimally in a narrow range of S expression. Over a range of 6.25 ng to 50 pg per well of S-encoding plasmid, none of the tested S mutations produced >twofold improvement in luciferase expression (fig. S4, A and B); slightly improved luciferase expression occurred with the S sequence derived from the Alpha variant (B.1.1.7) and with S containing the mutation N501Y within the receptor binding domain. These findings are in contrast to prior results for S-pseudotyped lentiviruses, for which enhanced entry was reported for some S mutations, including S:N501Y (*25*, *26*). However, S mutations tested in the context of SARS-CoV-2 infectious clones have shown mixed effects, and mutations within S may also mediate enhanced transmission by interfering with the binding of neutralizing antibodies (*27*, *28*). We tested whether VLPs could also be used to measure antibody neutralization and found results similar to previously reported data (Fig. 3C and fig. S4C). Using a neutralizing monoclonal antibody (MM43), we observed dose-dependent inhibition of luminescence with a measured median inhibitory concentration (IC_{50}) of 0.35 μg/ml, similar to the manufacturer-reported IC_{50} of 1.41 μg/ml. We also tested S from circulating variants and observed robust MM43-mediated neutralization of SC2-VLPs generated from Alpha, Beta, Gamma, B.1.617.2 (Delta), and Epsilon variant S proteins (Fig. 3D) and consistent with previous studies (*29*). These results show that SC2-VLPs employ S-mediated entry and can be used for screening S mutations for entry and neutralization. Although we did not observe enhanced entry resulting from the S mutations we tested, these mutations could still provide a fitness advantage for SARS-CoV-2 by limiting antibody-mediated neutralization. Detailed characterization of S mutations and their sensitivity to neutralizing antibodies has been examined in other studies (*30*, *31*).

We next used SC2-VLPs to test whether N mutations found in circulating variants improve viral particle assembly, RNA delivery, and/or reporter gene expression. We tested 15 N mutations, including two combinations corresponding to the Alpha and Gamma variants because they both contain the co-occurring R203K and G204R mutations. Alpha and Gamma variant N proteins improved luciferase expression in receiver cells by 7.5- and 4.2-fold, respectively, relative to the ancestral Wuhan Hu-1 N protein (Fig. 3E). In addition, four single–amino acid changes—P199L, S202R, R203K, and R203M—improved luciferase expression. Two of these mutations (P199L and R203K) do not change the overall charge, one (S202R) results in a more positive charge, and one (R203M) results in a more negative charge, which suggests that the improvement in luciferase expression is not likely due to simple electrostatics. Western blotting revealed no correlation between N protein expression levels and luciferase induction, suggesting that these N mutations enhance luciferase induction through a different mechanism (fig. S5). Notably, half of the amino acid changes observed within N and all of the mutations observed to enhance luminescence occur within a seven–amino acid span (199 to 205) of the central disordered region (termed the "linker" region; Fig. 3F), suggestive of a shared mechanism.

Further analysis of six N mutants was conducted to determine whether these mutations affect SC2-VLP assembly efficiency, RNA packaging, or RNA uncoating prior to expression. We chose the three mutants for which luciferase expression was improved by a factor of ~10 (P199L, S202R, and R203M) and two mutants that did not result in significantly increased luciferase expression (G204R and M234I) in the preliminary screen and repeated the measurement of their induced luciferase expression (Fig. 4A). N protein expression levels were similar in packaging cells, except for G204R, and did not correlate with luciferase expression (Fig. 4, A and B). We then purified SC2-VLPs containing each N mutation and found that those containing P199L and S202R had increased levels of S, N, and luciferase RNA, whereas R203M showed increased luciferase RNA only (Fig. 4C). These results suggest that mutations within the N linker domain improve the assembly of SC2-VLPs, leading to greater overall VLP production, a larger fraction of VLPs that contain RNA, or higher RNA content per particle. In each case, these results suggest a previously unanticipated explanation for the increased fitness and spread of SARS-CoV-2 variants of concern.

To validate whether the effects we observed in SC2-VLPs improve replication of intact virus, we used reverse genetics to generate SARS-CoV-2 containing N:S202R and N:R203M substitutions within a USA/WA1-2020 (Washington isolate) background (*1*) (Fig. 4D). We generated and used next-generation sequencing to verify stocks of virus containing the indicated mutations (fig. S6). We infected A549-ACE2 cells with wild type, N:S202R, or N:R203M at a multiplicity of infection (MOI) of 0.1 and collected supernatants at 24, 48, and 72 hours after infection. Reverse transcription quantitative polymerase chain reaction (RT-qPCR) indicated 45 and 23 times the RNA content in the supernatant at 72 hours after infection, and plaque assays indicated 166 and 51 times the infectious titers for N:S202R and N:R203M virus, respectively (Fig. 4, E to G). Our results indicate that both of these mutations enhance replication in lung epithelial cells, consistent with our observations using SC2-VLPs.

Overall, we present a strategy for rapidly generating and analyzing SC2-VLPs that package and deliver exogenous mRNA. This approach allows examination of viral assembly, budding, stability, maturation, entry, and genome uncoating involving all of the viral structural proteins (S, E, M, and N) without generating replication-competent virus. Such a strategy is useful not only for dissecting the molecular virology of SARS-CoV-2 but also for future development and screening of therapeutics to target assembly, budding, maturation, and entry. This strategy is ideally suited for the development of new antivirals targeting SARS-CoV-2, as it is sensitive, quantitative, and scalable to high-throughput workflows. The unexpected finding of improved mRNA packaging and luciferase induction by mutations within the N protein points to a previously unknown strategy for coronaviruses to evolve enhanced viral fitness. The mechanism for this enhancement involves increased mRNA packaging and delivery, although the exact process is currently unknown. Recent literature suggests that these mutations may affect phosphorylation of N, its binding to RNA, and

phase separation behavior—all of which could affect the efficiency of assembly (*32*–*34*). Our observations of enhanced RNA packaging and replication are consistent with recent reports that the Delta variant (containing N:R203M) generates viral RNA levels elevated by a factor of 1000 within patients (*35*). Our results provide a molecular basis to explain why the SARS-CoV-2 Delta variant demonstrates improved viral fitness.

REFERENCES AND NOTES

1. X. Xie *et al.*, *Cell Host Microbe* **27**, 841–848.e3 (2020).
2. S. Torii *et al.*, *Cell Rep.* **35**, 109014 (2021).
3. C. Ye *et al.*, *mBio* **11**, e02168-20 (2020).
4. X. Xie *et al.*, *Nat. Protoc.* **16**, 1761–1784 (2021).
5. S. J. Rihn *et al.*, *PLOS Biol.* **19**, e3001091 (2021).
6. J. A. Plante *et al.*, *Nature* **592**, 116–121 (2021).
7. K. H. D. Crawford *et al.*, *Viruses* **12**, 513 (2020).
8. A. A. Latif, Lineage comparison. outbreak.info (accessed 22 July 2021); https://outbreak.info/compare-lineages.
9. W. Zeng *et al.*, *Biochem. Biophys. Res. Commun.* **527**, 618–623 (2020).
10. J. Cubuk *et al.*, *Nat. Commun.* **12**, 1936 (2021).
11. T. M. Perdikari *et al.*, *EMBO J.* **39**, e106478 (2020).
12. C. B. Plescia *et al.*, *J. Biol. Chem.* **296**, 100103 (2021).
13. H. Swann *et al.*, *Sci. Rep.* **10**, 21877 (2020).
14. J. Lu *et al.*, *Cell Res.* **30**, 936–939 (2020).
15. Y. L. Siu *et al.*, *J. Virol.* **82**, 11318–11330 (2008).
16. P.-K. Hsieh *et al.*, *J. Virol.* **79**, 13848–13855 (2005).
17. S. Dent, B. W. Neuman, *Coronaviruses* **1282**, 99–108 (2015).
18. X. Lu *et al.*, *Immunology* **122**, 496–502 (2007).
19. L. Kuo, P. S. Masters, *J. Virol.* **87**, 5182–5192 (2013).
20. K. Woo, M. Joo, K. Narayanan, K. H. Kim, S. Makino, *J. Virol.* **71**, 824–827 (1997).
21. J. A. Fosmire, K. Hwang, S. Makino, *J. Virol.* **66**, 3522–3530 (1992).
22. A. Kanakan *et al.*, *Pathogens* **9**, 912 (2020).
23. T. Giroglou *et al.*, *J. Virol.* **78**, 9007–9015 (2004).
24. M. Ujike, C. Huang, K. Shirato, S. Makino, F. Taguchi, *J. Gen. Virol.* **97**, 1853–1864 (2016).
25. X. Deng *et al.*, *Cell* **184**, 3426–3437.e8 (2021).
26. A. Kuzmina *et al.*, *Cell Host Microbe* **29**, 522–528.e2 (2021).
27. Y. Liu *et al.*, *Nature* 10.1038/s41586-021-04245-0 (2021).
28. C. Motozono *et al.*, *Cell Host Microbe* **29**, 1124–1136.e11 (2021).
29. R. Kalkeri *et al.*, *Microorganisms* **9**, 1744 (2021).
30. Y. Weisblum *et al.*, *eLife* **9**, e61312 (2020).
31. W. T. Harvey *et al.*, *Nat. Rev. Microbiol.* **19**, 409–424 (2021).
32. X. Ju *et al.*, *PLOS Pathog.* **17**, e1009439 (2021).
33. C. R. Carlson *et al.*, *Mol. Cell* **80**, 1092–1103.e4 (2020).
34. S. Lu *et al.*, *Nat. Commun.* **12**, 502 (2021).
35. B. Li *et al.*, medRxiv 2021.07.07.21260122 [Preprint] (2021); doi:10.1101/2021.07.07.21260122.

ACKNOWLEDGMENTS

We thank P.-Y. Shi for providing the plasmids for SARS-CoV-2 reverse genetics. We would also like to acknowledge J. Hamilton for advice and guidance and S. Wyman and B. Bach for sequencing of infectious clones. We also thank S. Pillai and H. S. Sperbert at Vitalant Research Institute (San Francisco, CA) for providing the 293T-ACE2-TMPRSS2 cells used for pseudovirus entry assays, the S. P. J. Whelan laboratory for providing Vero cells overexpressing human TMPRSS2, O. Schwartz from Institut Pasteur for providing A549 cells stably expressing ACE2 (A549-ACE2), and K.-K. Conzelmann for providing BSR-T7/5 cells. **Funding:** This project was funded by a grant from the National Institutes of Health (NIH; R21AI59666). J.A.D. is a Howard Hughes Medical Institute investigator. A.M.S. acknowledges the support of the Natural Sciences and Engineering Research Council of Canada (NSERC PDF-533021-2019) and IPC support from the NIH (F31 AI164671-01). **Author contributions:** Conceptualization: A.M.S. and J.A.D. Investigation: A.M.S., T.Y.T., T.T., I.P.C., A.C., M.M.K., B.S., P.-Y.C., J.M.H., and K.M.S. Methodology: A.M.S., T.Y.T., T.T., J.A.D., and M.O. Supervision: J.A.D. and M.O. Writing: A.M.S., T.Y.T., T.T., and J.A.D. **Competing interests:** A.M.S. and J.A.D. are inventors on a patent application filed by the Gladstone Institutes and the University of California that covers the method and composition of SARS-CoV-2 VLP preparations for RNA transduction and expression in cells. J.A.D. is a cofounder of Caribou Biosciences, Editas Medicine, Scribe Therapeutics, Intellia Therapeutics, and Mammoth Biosciences. J.A.D. is a scientific advisory board member of Vertex, Caribou Biosciences, Intellia Therapeutics, eFFECTOR Therapeutics, Scribe Therapeutics, Mammoth Biosciences, Synthego, Algen Biotechnologies, Felix Biosciences, The Column Group, and Inari. J.A.D. is a director at Johnson & Johnson and Tempus and has research projects sponsored by Biogen, Pfizer, AppleTree Partners, and Roche. **Data and materials availability:** All data are available in the main paper or the supplementary materials. Plasmids are available from Addgene (addgene.org) or by request. 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.org/doi/10.1126/science.abl6184
Materials and Methods
Figs. S1 to S6
Tables S1 to S3
References
MDAR Reproducibility Checklist

26 July 2021; accepted 29 October 2021
Published online 4 November 2021
10.1126/science.abl6184

IMMUNOTHERAPY

Dietary fiber and probiotics influence the gut microbiome and melanoma immunotherapy response

Christine N. Spencer[1]†‡, Jennifer L. McQuade[2]†, Vancheswaran Gopalakrishnan[1]†§, John A. McCulloch[3]†, Marie Vetizou[3]†, Alexandria P. Cogdill[1,4]†¶, Md A. Wadud Khan[1], Xiaotao Zhang[5], Michael G. White[1], Christine B. Peterson[6], Matthew C. Wong[1], Golnaz Morad[1], Theresa Rodgers[2], Jonathan H. Badger[3], Beth A. Helmink[1]#, Miles C. Andrews[1,7], Richard R. Rodrigues[8], Andrey Morgun[9], Young S. Kim[10], Jason Roszik[2], Kristi L. Hoffman[11], Jiali Zheng[5], Yifan Zhou[4], Yusra B. Medik[4], Laura M. Kahn[4,12], Sarah Johnson[1], Courtney W. Hudgens[13], Khalida Wani[13], Pierre-Olivier Gaudreau[14], Angela L. Harris[15], Mohamed A. Jamal[16], Erez N. Baruch[17], Eva Perez-Guijarro[18], Chi-Ping Day[18], Glenn Merlino[18], Barbara Pazdrak[2], Brooke S. Lochmann[2], Robert A. Szczepaniak-Sloane[1], Reetakshi Arora[1], Jaime Anderson[2], Chrystia M. Zobniw[2], Eliza Posada[2], Elizabeth Sirmans[2], Julie Simon[1], Lauren E. Haydu[1], Elizabeth M. Burton[1], Linghua Wang[16], Minghao Dang[16], Karen Clise-Dwyer[19,20], Sarah Schneider[19], Thomas Chapman[1], Nana-Ama A. S. Anang[4], Sheila Duncan[1], Joseph Toker[21,22], Jared C. Malke[1], Isabella C. Glitza[2], Rodabe N. Amaria[2], Hussein A. Tawbi[2], Adi Diab[2], Michael K. Wong[2], Sapna P. Patel[2], Scott E. Woodman[2], Michael A. Davies[2], Merrick I. Ross[1], Jeffrey E. Gershenwald[1], Jeffrey E. Lee[1], Patrick Hwu[2]††, Vanessa Jensen[23], Yardena Samuels[24], Ravid Straussman[24], Nadim J. Ajami[16], Kelly C. Nelson[25], Luigi Nezi[26], Joseph F. Petrosino[11], P. Andrew Futreal[16], Alexander J. Lazar[12,16,27], Jianhua Hu[28], Robert R. Jenq[16,29], Michael T. Tetzlaff[30], Yan Yan[31], Wendy S. Garrett[32], Curtis Huttenhower[31,33,34,35], Padmanee Sharma[4,36,37], Stephanie S. Watowich[4], James P. Allison[4,37], Lorenzo Cohen[38]‡‡, Giorgio Trinchieri[3]*‡‡, Carrie R. Daniel[5]*‡‡, Jennifer A. Wargo[1,16]*‡‡**

Gut bacteria modulate the response to immune checkpoint blockade (ICB) treatment in cancer, but the effect of diet and supplements on this interaction is not well studied. We assessed fecal microbiota profiles, dietary habits, and commercially available probiotic supplement use in melanoma patients and performed parallel preclinical studies. Higher dietary fiber was associated with significantly improved progression-free survival in 128 patients on ICB, with the most pronounced benefit observed in patients with sufficient dietary fiber intake and no probiotic use. Findings were recapitulated in preclinical models, which demonstrated impaired treatment response to anti–programmed cell death 1 (anti–PD-1)–based therapy in mice receiving a low-fiber diet or probiotics, with a lower frequency of interferon-γ–positive cytotoxic T cells in the tumor microenvironment. Together, these data have clinical implications for patients receiving ICB for cancer.

Treatment with immune checkpoint blockade (ICB) has revolutionized cancer therapy (*1*), and the influence of the gut microbiome on therapeutic response has now been demonstrated in numerous human cohorts and in preclinical models (*2*–*7*). The human gut microbiome is itself shaped by a wide variety of environmental exposures, including diet (*8*, *9*) and medication use (*10*–*13*), with host genetics accounting for <10% of variation (*14*). However, whether factors such as dietary fiber intake and the use of commercially available probiotics affect immunotherapy responses in cancer patients remains unclear.

To help address this, we profiled the gut (fecal) microbiome and assessed clinicopathologic features and outcomes in a large cohort of melanoma patients (n = 438; Fig. 1A and fig. S1). The majority of these patients were receiving systemic therapy for metastatic melanoma (n = 321), and responses to treatment were assessed with radiographic imaging

in those with evaluable treatment responses (n = 293), classifying patients as either responders [(R) complete or partial response or stable disease ≥6 months; n = 193] or non-responders [(NR) stable disease <6 months or progressive disease; n = 100] using Response Evaluation Criteria in Solid Tumors (RECIST 1.1) (*15*). The majority of patients were treated with ICB (87%), most commonly anti–programmed cell death 1 (anti–PD-1) therapy (Fig. 1A, fig. S1, and table S1). Patients initiating therapy with ICB were asked to co-enroll onto a lifestyle survey protocol, which included baseline assessments of dietary habits and use of probiotic supplements within the past month (n = 158; Fig. 1A, fig. S1, and table S1) (*16*, *17*).

We first assessed the relative abundance of gut microbial taxa associated with response to anti–PD-1 immunotherapy in our prior published study (*4*) within a larger cohort of newly accrued anti–PD-1–treated patients (n = 132 total; n = 87 R and n = 45 NR), excluding patients from the previously published cohort. On the basis of our prior study, we hypothesized that bacteria from the Ruminococcaceae family and *Faecalibacterium* genus would be associated with response to therapy. We tested this by specifically querying the abundance of these taxa in responders versus nonresponders to anti–PD-1, again observing enrichment of both taxa (Fig. 1B) as well as of *Faecalibacterium prausnitzii* in the metagenomic subset (n = 111 total; n = 71 R and n = 40 NR; fig. S2A) in anti–PD-1 responders. We did not observe significant differences in the alpha and beta diversity of the gut microbiota in responders versus nonresponders (fig. S2, E and F), in contrast to our prior study. This discrepancy may reflect associations driven by a small number of patients in the prior study with improved power and reduced error in the larger cohort (fig. S3 and tables S2 to S4), and it underscores the lack of concordance across numerous studies that implicate gut bacteria in response to cancer immunotherapy (*18*).

Next, we assessed the composition of the gut microbiome in responders and nonresponders in the full cohort of late-stage melanoma patients with evaluable responses to any systemic therapy (n = 293 total; n = 193 R and n = 100 NR; Fig. 1C), as well as in all patients treated with anti–PD-1 monotherapy (fig. S2), including patients from both the newly accrued and the previously published cohorts. Across the full cohort, we observed a significantly higher abundance of Ruminococcaceae in the gut microbiota of responders versus nonresponders treated with anti–PD-1 or other systemic therapies that remained consistent after adjustment for potential confounders [age, sex, body mass index (BMI), prior treatment, and antibiotic use] (Fig. 1C, fig. S2, and tables S5 to S9). However, we did not observe significant differences in the overall composition of the gut microbiota in responders versus nonresponders in this larger cohort of patients on systemic therapy (fig. S2 and table S9), nor was there strong concordance with response-associated taxa from the prior study—beyond Ruminococcaceae—in the newly accrued cohort (figs. S3 and S4 and tables S3 and S10). We also assessed the abundance of our previously reported response-associated taxa in published datasets from two recently completed clinical trials demonstrating potential efficacy of the use of fecal microbiota transplant (FMT) + anti–PD-1 in immunotherapy-refractory melanoma patients (*19*, *20*), noting that many of our response-associated taxa appeared to be enriched in the post-FMT specimens from patients who responded to this treatment (Fig. 1, D and E, and fig. S5).

Given that cancer patients are increasingly interested in using probiotic supplements to augment gut health, we assessed the use of commercially available probiotics within our cohort and observed that 31% (49 of 158) of late-stage melanoma patients initiating ICB reported that they had taken a probiotic supplement within the past month. Patients who reported taking a probiotic supplement preceding the start of treatment with ICB had a lower BMI, were less likely to take statins, and reported slightly higher intake of vegetables and legumes than patients who did not take probiotic supplements (table S1). The proportion of patients reporting antibiotic use within the past month was markedly similar in those who did (29%) versus did not (28%) report probiotic use. Steroid or proton-pump inhibitor use was also not associated with probiotic use (table S1). We then assessed whether probiotic use was associated with differential outcomes in patients treated with ICB and observed no statistically significant differences in progression-free survival (PFS) (n = 158; Fig. 2A; median PFS 17 versus 23 months; Table 1) or odds of response in patients who reported taking probiotics (59% R) versus those who did not (68% R) (Table 1 and tables S1, S11, and S12). The modest associations of probiotic use and outcomes in this cohort were not surprising to us because limitations existed regarding the overall cohort size, as well as substantial heterogeneity in

[1]Department of Surgical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. [2]Department of Melanoma Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. [3]Laboratory of Integrative Cancer Immunology, Center for Cancer Research (CCR), National Cancer Institute (NCI), National Institutes of Health (NIH), Bethesda, MD 20892, USA. [4]Department of Immunology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. [5]Department of Epidemiology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. [6]Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. [7]Department of Medicine, Monash University, Melbourne, VIC 3004, Australia. [8]Frederick National Laboratory for Cancer Research, and Microbiome and Genetics Core, Laboratory of Integrative Cancer Immunology, CCR, NCI, NIH, Bethesda, MD 20852, USA. [9]Department of Pharmaceutical Science, Oregon State University, Corvallis, OR 97331, USA. [10]Nutritional Science Research Group, Division of Cancer Prevention, NCI, NIH, Rockville, MD 20850, USA. [11]Department of Molecular Virology and Microbiology, Baylor College of Medicine, Houston, TX 77030, USA. [12]MD Anderson University of Texas Health Graduate School, Houston, TX 77030, USA. [13]Department of Translational Molecular Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. [14]Canadian Cancer Trials Group and Department of Oncology, Queen's University, Kingston, ON K7L 3N6, Canada. [15]Center for Co-Clinical Trials, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. [16]Department of Genomic Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. [17]Department of Internal Medicine, The University of Texas Health Science Center, Houston, TX 77030, USA. [18]Laboratory of Cancer Biology and Genetics, CCR, NCI, NIH, Bethesda, MD 20892, USA. [19]Advanced Cytometry and Sorting Facility at South Campus, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. [20]Department of Hematopoietic Biology and Malignancy, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. [21]Department of Neurosurgery, Harvard University, Cambridge, MA 02138, USA. [22]Department of Oncology, University of Cambridge, Cambridge CB2 1TN, UK. [23]Department of Veterinary Medicine and Surgery, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. [24]Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot, 76100, Israel. [25]Department of Dermatology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. [26]Dipartimento di Oncologia Sperimentale, Instituto Europeo di Oncologia, Milan, P.I. 08691440153, Italy. [27]Department of Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. [28]Department of Biostatistics, Columbia University, New York, NY 10032, USA. [29]Department of Stem Cell Transplant, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. [30]Departments of Pathology and Dermatology, Dermatopathology and Oral Pathology Unit, University of California San Francisco, San Francisco, CA 94115, USA. [31]Department of Biostatistics and the Harvard T.H. Chan Microbiome in Public Health Center, Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA 02115, USA. [32]Department of Molecular Metabolism, T.H. Chan School of Public Health, Harvard University, Boston, MA 02115, USA. [33]Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA. [34]Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA. [35]Harvard Chan Microbiome in Public Health Center, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA. [36]Department of Genitourinary Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. [37]Parker Institute for Cancer Immunotherapy, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA. [38]Department of Palliative, Rehabilitation, and Integrative Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA.

***Corresponding author. Email: jwargo@mdanderson.org (J.A.W.); cdaniel@mdanderson.org (C.R.D.); trinchig@niaid.nih.gov (G.T.)**

†These authors contributed equally to this work.

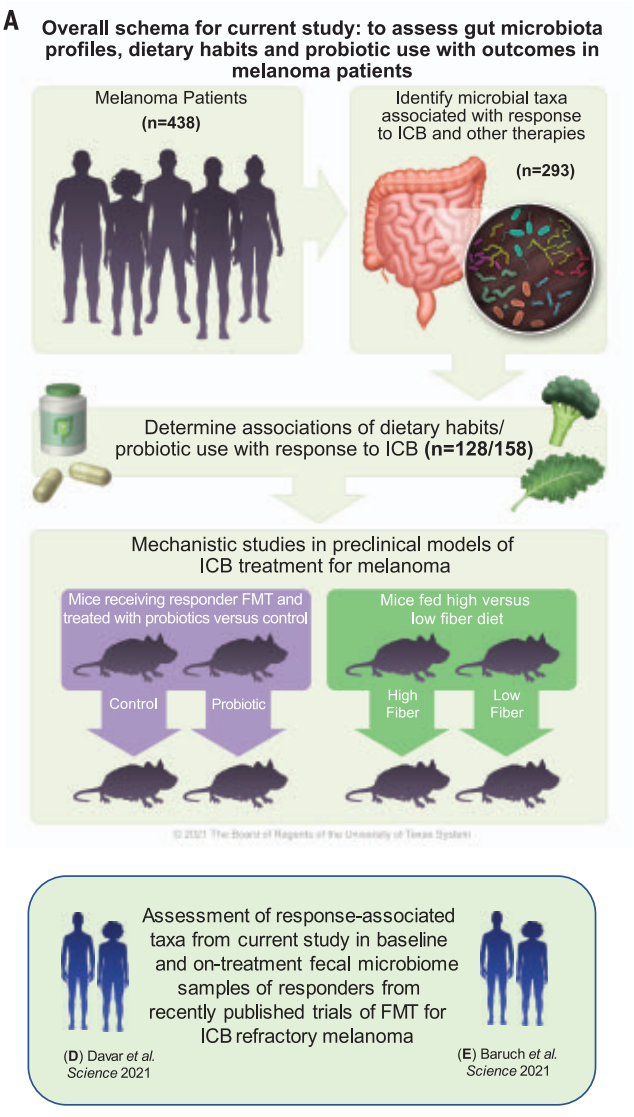
‡Present address: Department of Informatics, Parker Institute for Cancer Immunotherapy, San Francisco, CA 94129, USA.

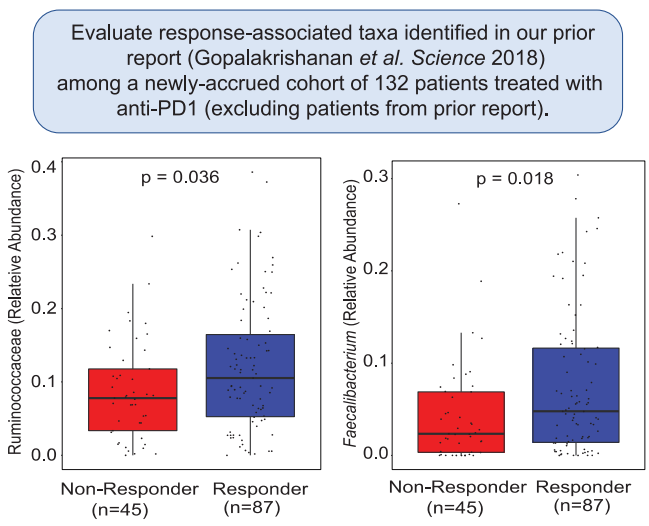
§Present address: AstraZeneca, Gaithersburg, MD 20878, USA.

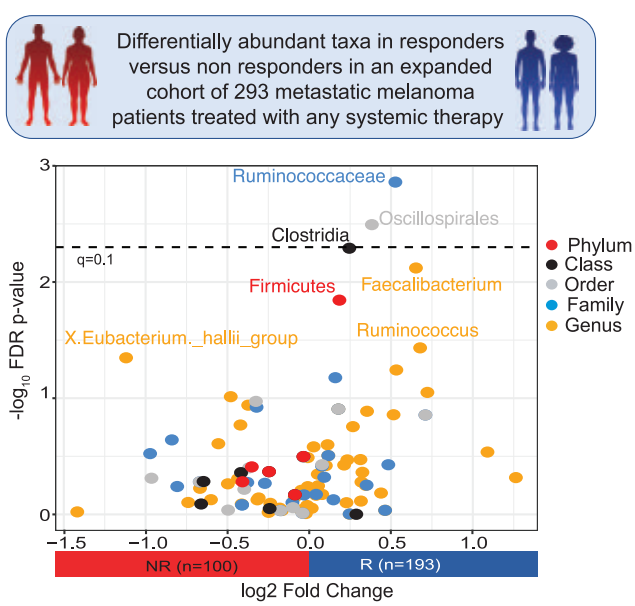
¶Present address: Immunai, New York, NY 10013, USA.

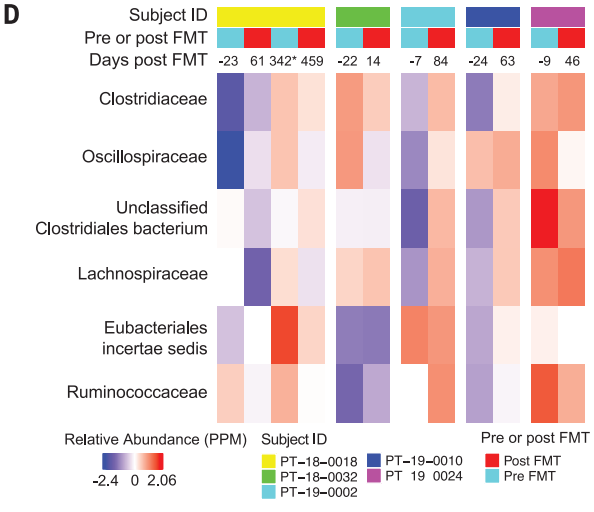
#Present address: Department of Surgery, Washington University School of Medicine, St. Louis, MO 63110, USA.

**Present address: Department of Epidemiology and Biostatistics, School of Public Health, Shanghai Jiao Tong University School of Medicine, Shanghai 200025, China. ††Present address: Administration, Moffitt Cancer Center, Tampa, FL 33612, USA. ‡‡These authors contributed equally to this work.









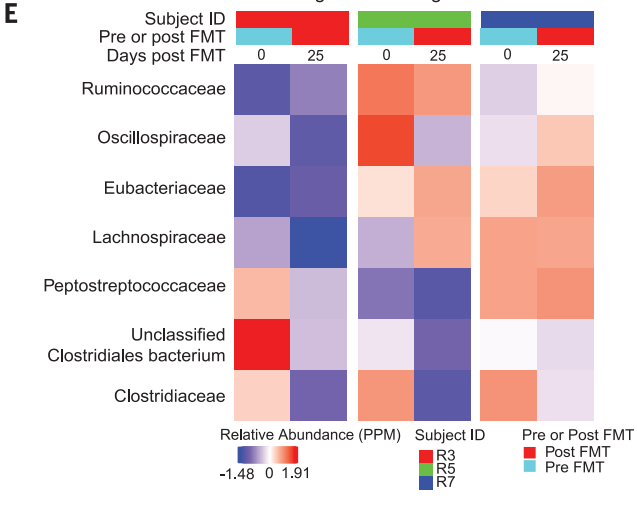


Fig. 1. Profiles of gut microbiota in patients with melanoma and associations with outcomes on therapy. (**A**) Schema of study design. (**B**) Box plots comparing the relative abundance of anti–PD-1 response–associated taxa from Gopalakrishnan *et al.* (*4*) with a newly recruited cohort (n = 132) of anti–PD-1–treated patients (P = 0.036 and P = 0.018, respectively, for Ruminococcaceae and *Faecalibacterium* by Wilcoxon rank sum test). Patients included in the prior

study were excluded from this analysis. (**C**) Volcano plot depicting pairwise comparisons of relative abundances of bacterial taxa. The *y* axis displays the $-\log_{10}$ false discovery rate (FDR)–corrected *P* value (dashed line, q < 0.1), and the *x* axis shows the log2 fold change comparing 193 R and 100 NR patients with systemic therapy across the full cohort, including patients from the prior study (by Wilcoxon rank sum test with FDR correction per level). (**D**) Heatmap of scaled relative abundances [parts per million (PPM)] of bacteria belonging to order Clostridiales and family Ruminococcaceae in pre- and post-FMT samples of anti–PD-1 refractory metastatic melanoma FMT recipients who responded to FMT + anti–PD-1 in Davar *et al*. (*20*) [National Center for Biotechnology Information (NCBI) accession no. PRJNA672867]. Number of days from FMT are depicted on the top of each heatmap column, with post-FMT values being the geometric mean days of all post-FMT time points for that patient. The geometric mean of relative abundances of post-FMT samples from each patient were calculated as the single post-FMT mean relative abundance. The exception is patient PT–18–0018, who received two FMTs (denoted by an asterisk). The first post-FMT column for this patient reflects the geometric mean of samples leading up to the second FMT event. (**E**) Heatmap of scaled relative abundances of bacteria belonging to order Clostridiales and family Ruminococcaceae in pre- and post-FMT samples of anti–PD-1 refractory metastatic melanoma FMT recipients who responded to FMT + anti–PD-1 in Baruch *et al*. (*19*) (NCBI accession no. PRJNA678737). Number of days from FMT are depicted on the top of each heatmap column, with post-FMT values being the geometric mean days of all post-FMT time points for that patient. The geometric mean of relative abundances of post-FMT samples from each patient were calculated as the single post-FMT mean relative abundance.

the specific probiotic supplement(s) reportedly used by patients. Although we did not observe statistically significant differences in outcomes or in microbiota features (fig. S6 and table S9) in patients on ICB by probiotic use, the overall trends observed were intriguing—particularly given the relatively high proportion of patients reporting probiotic supplementation in this cohort. Thus, we sought to examine the effects of probiotic use on response to ICB in preclinical models.

To do this, germ-free mice first received FMT using donor stool from a complete responder (CR) patient to anti–PD-1 blockade. After this, mice were orally gavaged with one of two commercially purchased probiotics (*Bifidobacterium longum*– or *Lactobacillus rhamnosus* GG–based) versus sterile water control. Viability and composition of the bacterial strains in the probiotic were confirmed by culture and sequencing (fig. S7). Mice were then challenged with murine melanoma tumors and treated with anti–PD-1 ligand 1 (anti–PD-L1) therapy (because treatment with this antibody is more effective in this particular murine tumor model than anti–PD-1) (Fig. 2B). In these studies, mice receiving probiotics demonstrated impaired antitumor response to treatment with anti–PD-L1 and had significantly larger tumors compared with control mice (Fig. 2C), with findings that were recapitulated in an additional murine tumor model (fig. S8). Notably, similar findings were also observed in non–germ-free and specific pathogen–free (SPF) mice implanted with melanoma tumors (fig. S8) that harbor a microbiota from birth to which they are well coadapted.

We next compared the gut microbiota of mice receiving probiotics versus sterile water control, and we observed differences in gut microbiota diversity in the mice receiving probiotics compared with control (Fig. 2, D and E, and fig. S9). Analysis of tumor-infiltrating immune subsets from anti–PD-L1–treated mice revealed a significantly reduced frequency of interferon-γ (IFN-γ) positive $CD8^+$ T cells in tumors of probiotic-treated mice versus controls (Fig. 2F). A trend toward fewer IFN-γ $CD4^+$ T helper 1 (T_H1) cells in tumors from mice receiving probiotics versus control was also observed, although this did not reach statistical significance (Fig. 2G). Unsupervised analyses of the flow cytometry data corroborated the findings in immune subsets between probiotic treatment versus control, demonstrating a reduced frequency of cytotoxic T cells in the tumor microenvironment of probiotic-treated mice (Fig. 2H). These data are in line with previously published studies that have demonstrated increased tumorigenesis in murine models of colorectal carcinoma in probiotic-treated mice (*21*), although other studies have shown a beneficial effect of other probiotic formulations and rationally designed bacterial consortia in preclinical models and patient cohorts (*22*–*24*). Together, these studies support the need for more careful investigations of the effects of current commercially available probiotic formulations on immunity and cancer immunotherapy response.

Given that many of the response-associated bacteria identified in our cohort have known roles in starch degradation and fiber fermentation (*25*–*29*), we next sought to assess the effect of dietary fiber intake on response to ICB. We asked patients who were initiating treatment with ICB to complete the National Cancer Institute Dietary Screener Questionnaire (NCI-DSQ) (*17*), and responses were scored to derive dietary fiber intake from 26 queried food items. Dietary fiber intake was assessed per 5-g/day incremental increase and further categorized according to the distribution of reported intake within our cohort with low or insufficient fiber intake corresponding to <20 g/day and sufficiently high fiber intake at or above 20 g/day, a threshold met by ~30% (37 of 128) of ICB patients (Fig. 3A and fig. S10A). As expected, dietary fiber intake was highly correlated with fruit, vegetable, legume, and whole grain intake and, to a lesser extent, with calcium intake (fig. S10B and table S1). Patients with insufficient dietary fiber intake were more likely to be obese—a factor that we and others have previously found to be paradoxically associated with improved response to ICB (*30*, *31*)—and were also more likely to take antihypertensive medications (table S1).

Patients who reported sufficient dietary fiber intake (n = 37 of 128) demonstrated improved PFS over those with insufficient dietary fiber intake (median PFS not reached versus 13 months; Fig. 3A and Table 1). After adjustment for clinical factors, every 5-g increase in daily dietary fiber intake corresponded with a 30% lower risk of progression or death (Table 1). Similar associations were observed when assessing dietary fiber intake in relation to the odds of response to ICB (Table 1). The observed protective effect of dietary fiber intake in relation to PFS and response remained consistent among the subset of patients treated with anti–PD-1 monotherapy, with the exclusion of patients reporting recent antibiotic use given the known impact of these on ICB response (*32*) (tables S11 and S12). We did not observe substantial differences in the composition of the gut microbiota in those who reported sufficient versus insufficient fiber intake as assessed by 16*S* and metagenomic sequencing (fig. S11 and table S9); however, this is not unexpected given the known challenges of isolating associations of specific dietary components from other factors known to affect the gut microbiota in observational human cohorts (*33*, *34*).

After this, we further evaluated whether dietary fiber intake and probiotic use may jointly affect clinical outcomes in patients treated with ICB, given the potential associations between these factors. In this cohort, patients with sufficient dietary fiber intake were somewhat more likely to take probiotics than those reporting insufficient dietary fiber intake (35 versus 27%; table S1). We assessed potential additive effects across a combined variable comparing four groups of patients—including those reporting insufficient dietary fiber intake with no probiotic use (53%), those reporting insufficient dietary fiber intake with probiotic use (19%), those reporting sufficient dietary fiber intake with no probiotic use (18%), and those reporting sufficient dietary fiber intake with probiotic use (10%) (n = 123 patients total; Fig. 3B). Differences in outcomes were noted across the groups (Fig. 3B), with significantly longer PFS observed in patients

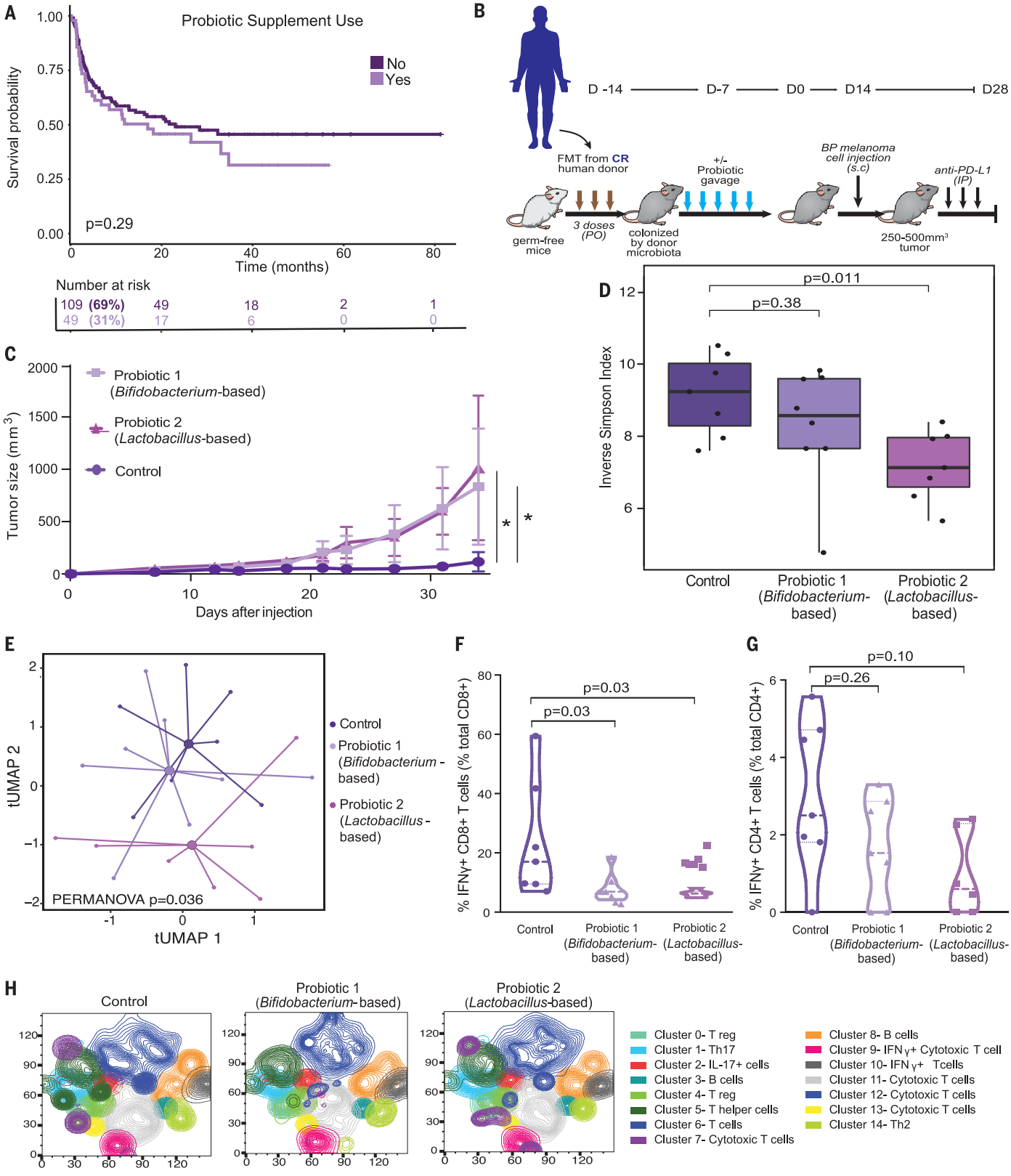


Fig. 2. Effect of probiotic supplement use in patients and in preclinical models of melanoma immunotherapy. (**A**) Kaplan-Meier plot comparing progression-free survival intervals by probiotic use among patients who received ICB (n = 158; P = 0.29 by log-rank test). (**B**) Experimental design of studies in germ-free (GF) mice that received FMT from a complete responder (CR) donor combined with probiotic 1, probiotic 2, or sterile water control before tumor injection [2.5 × 10^5 to 8 × 10^5 $BRAF^{V600E}/PTEN^{-/-}$ (BP) tumor cells] and treatment with anti–PD-L1. Time is in days relative to tumor injection [day 0 (D0)]. PO, per orem; s.c, subcutaneous; IP, intraperitoneal. (**C**) Mouse tumor growth curves comparing volume of tumors in mice who received probiotics or

sterile water control (n = 4 to 5 per group); probiotic 1 versus probiotic 2 versus sterile water control. Data are means ± SEM tumor volume. All P values are from a likelihood ratio test in a linear mixed model (P = 0.04 *Bifidobacterium longum* 35624–based probiotic 1 versus control; P = 0.01 *Lactobacillus rhamnosus* GG–based probiotic 2 versus control). *P < 0.05. (**D**) Box plots comparing alpha diversity of the gut microbiome, as measured by the inverse Simpson index in mice treated with control, probiotic 1 (*Bifidobacterium longum* 35624–based), or probiotic 2 (*Lactobacillus rhamnosus* GG–based) (pairwise P values compared with control were calculated by Wilcoxon rank sum test). Fecal samples were collected for microbiome analysis (via metagenomic sequencing) from tumor-bearing mice before the anti–PD-L1 therapy (n = 7 to 8 per group), mimicking baseline sample collection from patients. (**E**) Ordination plot by t-distributed uniform manifold approximation and projection (t-UMAP) by Bray-Curtis distance, demonstrating compositional differences of the gut microbiome in mice treated with sterile water control, probiotic 1 (*Bifidobacterium longum* 35624–based), or probiotic 2 (*Lactobacillus rhamnosus* GG–based) [permutational multivariate analysis of variance (PERMANOVA) P = 0.036]. (**F** and **G**) Pairwise comparisons of sterile water control versus probiotic 1 (*Bifidobacterium longum* 35624–based) or control versus probiotic 2 (*Lactobacillus rhamnosus* GG–based) groups (n = 6 per group) via supervised analysis with manual gating for either frequency of IFN-γ^+ $CD8^+$ T cells in tumors (percent total tumor $CD8^+$ T cells) (P = 0.03, P = 0.03) (F) or frequency of IFN-γ^+ $CD4^+$ T cells in tumors (percent total tumor $CD4^+$ T cells) (P = 0.26, P = 0.10) (G). (**H**) Unsupervised analysis of flow cytometry data showing density t-distributed stochastic neighbor embedding (t-SNE) plot of tumor-infiltrating immune cells overlaid with color-coded clusters, with an equal number of $CD45^+$ infiltrating leukocytes for each treatment group (control, probiotic 1, and probiotic 2).

Table 1. Associations of baseline probiotic supplement use and dietary fiber intake in late-stage melanoma patients treated with ICB and followed for tumor response and progression-free survival. Dashes indicate not applicable. HR, hazard ratio; CI, confidence interval; OR, odds ratio; N/R, not reached; ref, referent group.

Comparison	*n*	Progression-free survival					Odds of response to ICB			
		Events	Median months	HR*	95% CI	*P* value†	Responder *n* (%)	OR*	95% CI	*P* value†
Probiotic supplement use										
Total	158	85	–	–	–	–	65%	–	–	–
No	109	56	23	1.00	ref	–	74 (68%)	1.00	ref	–
Yes	49	29	17	1.30	0.82, 2.07	0.27	29 (59%)	0.79	0.37, 1.66	0.52
Dietary fiber intake										
Total	128	73	–	–	–	–	65%	–	–	–
Per 5 g/day increase	–	–	–	0.71	0.52, 0.98	0.04	–	1.70	0.97, 3.00	0.06
Insufficient	91	57	13	1.00	ref	–	55 (60%)	1.00	ref	–
Sufficient	37	16	N/R	0.59	0.33, 1.04	0.07	28 (76%)	2.20	0.86, 5.61	0.10
Dietary fiber intake + probiotic supplement use										
Total	123	72	–	–	–	–	63%	–	–	–
Sufficient fiber + no probiotics	22	8	N/R	0.44	0.21, 0.92	0.03	18 (82%)	2.94	0.87, 9.94	0.08
Other‡	101	64	13	1.00	ref	–	60 (59%)	1.00	ref	–

*HR and 95% CI estimated using Cox proportional hazards regression. OR and 95% CI estimated using logistic regression. All models include multivariable adjustment for subtype, stage, lactate dehydrogenase level, and BMI. †*P* value by Wald test. ‡Other category includes patients who either reported insufficient fiber intake or probiotic use.

reporting sufficient dietary fiber intake and no probiotic use compared with all other groups (median PFS not reached versus 13 months; Fig. 3B and Table 1). Similar positive associations were observed for ICB response in patients reporting sufficient dietary fiber intake and no probiotic use compared with all other groups (n = 123; 82 versus 59% responders; Table 1 and tables S11 and S12). Microbial alpha diversity and Ruminococcaceae family and *Faecalibacterium* genus abundances were also numerically higher in patients with sufficient dietary fiber intake and no probiotic use, although only 18% of patients met these criteria and results did not reach statistical significance (fig. S12).

Intrigued by these findings, we next examined whether dietary fiber modulation could enhance therapeutic response to ICB in preclinical melanoma models. In these studies, conventionally housed C57BL/6 SPF mice were provided with a standard fiber-rich whole grain diet (17.6% fiber) versus a fiber-poor diet (2% fiber) (*29*), challenged with murine melanoma tumors (*35*), and treated with anti–PD-1 therapy versus isotype control (Fig. 3C). Mice receiving a fiber-rich diet demonstrated delayed tumor outgrowth compared with mice who received a fiber-poor diet when treated with anti–PD-1 (Fig. 3D). These findings were recapitulated in additional tumor models (figs. S13 and S14). By contrast, there was no effect of fiber-rich versus fiber-poor diet on the response to anti–PD-1 therapy in germ-free mice, which supports the hypothesis that the effect of this dietary intervention on treatment efficacy is microbiota dependent (fig. S13). Profiling of the gut microbiome revealed significant differences in the community structure of mice fed fiber-rich versus fiber-poor diets (Fig. 3E) and taxonomic differences between the groups (fig. S15).

Stool metabolomic profiling also revealed significantly higher levels of the short chain fatty acid (SCFA) propionate in mice receiving a fiber-rich diet, although no significant differences were noted in SCFA levels as a whole (fig. S16). Immune profiling by flow cytometry of tumors in treated mice revealed a significantly higher frequency of $CD4^+$ T cells overall (and those expressing PD-1) in the tumors of mice on high- versus low-fiber diets (fig. S17, A and B). We next conducted RNA sequencing of $CD45^+$ tumor-infiltrating lymphocytes (TILs) and observed significantly higher expression of genes related to T cell activation and interferon response in mice receiving a high- versus low-fiber diet in the setting of treatment with anti–PD-1 (Fig. 3, F and G, and tables S13 and S14). Further, network analysis of murine data suggested that the fiber-fermenting Ruminococcaceae family of bacteria may contribute to the effects

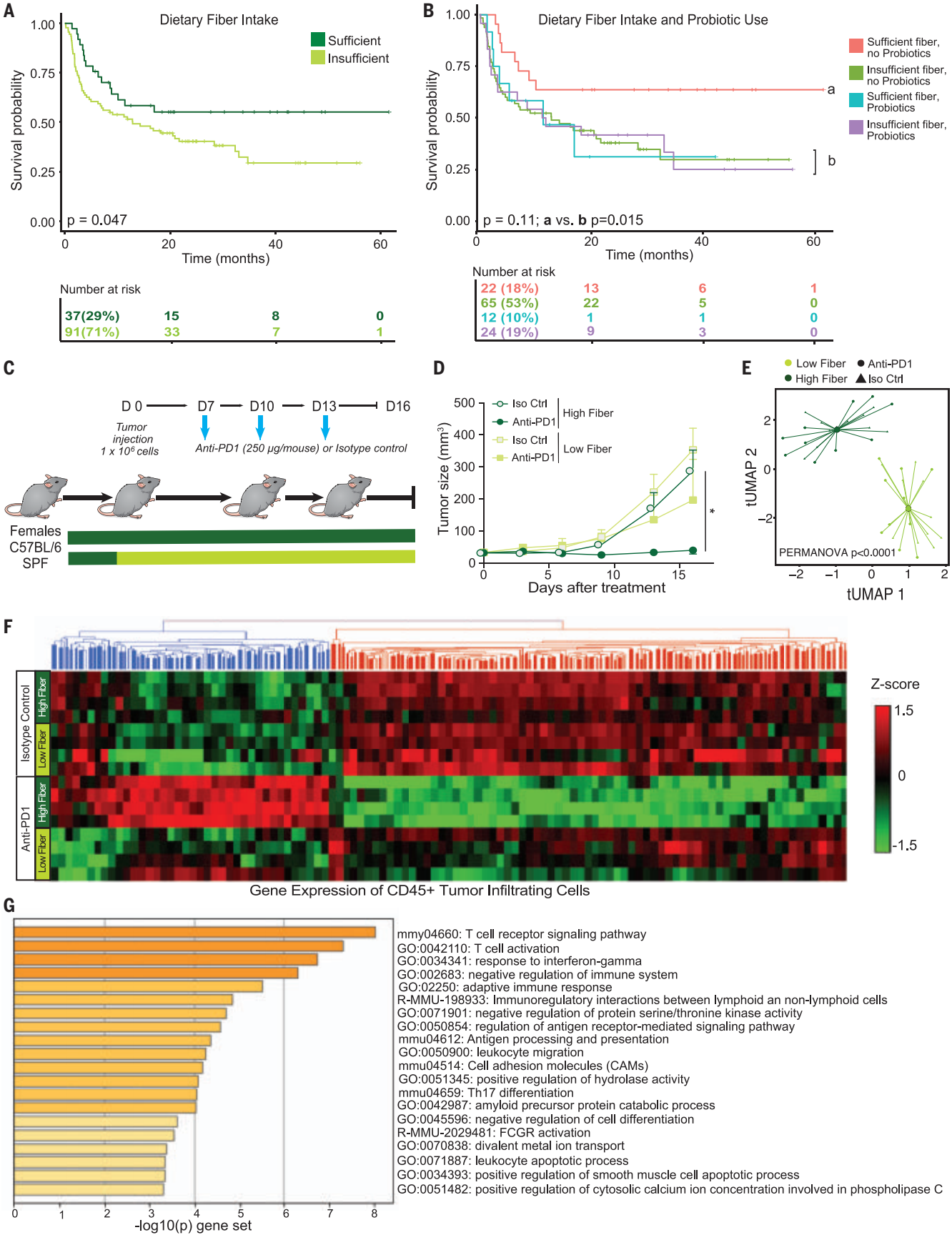

A
Dietary Fiber Intake
Sufficient
Insufficient
Survival probability
1.00
0.75
0.50
0.25
0.00
p = 0.047
0
20
40
60
Time (months)
Number at risk
37(29%) 15 8 0
91(71%) 33 7 1
B
Dietary Fiber Intake and Probiotic Use
Sufficient fiber, no Probiotics
Insufficient fiber, no Probiotics
Sufficient fiber, Probiotics
Insufficient fiber, Probiotics
a
b
p = 0.11; a vs. b p=0.015
Number at risk
22 (18%) 13 6 1
65 (53%) 22 5 0
12 (10%) 1 1 0
24 (19%) 9 3 0
C
D 0
D7
D10
D13
D16
Tumor injection 1 x 10^6 cells
Anti-PD1 (250 µg/mouse) or Isotype control
Females C57BL/6 SPF
D
Iso Ctrl
Anti-PD1
High Fiber
Low Fiber
Tumor size (mm^3)
Days after treatment
*
E
Low Fiber
High Fiber
Anti-PD1
Iso Ctrl
tUMAP 2
tUMAP 1
PERMANOVA p<0.0001
F
Isotype Control
Anti-PD1
High Fiber
Low Fiber
Z-score
1.5
0
-1.5
Gene Expression of CD45+ Tumor Infiltrating Cells
G
-log10(p) gene set
mmy04660: T cell receptor signaling pathway
GO:0042110: T cell activation
GO:0034341: response to interferon-gamma
GO:002683: negative regulation of immune system
GO:02250: adaptive immune response
R-MMU-198933: Immunoregulatory interactions between lymphoid an non-lymphoid cells
GO:0071901: negative regulation of protein serine/thronine kinase activity
GO:0050854: regulation of antigen receptor-mediated signaling pathway
mmu04612: Antigen processing and presentation
GO:0050900: leukocyte migration
mmu04514: Cell adhesion molecules (CAMs)
GO:0051345: positive regulation of hydrolase activity
mmu04659: Th17 differentiation
GO:0042987: amyloid precursor protein catabolic process
GO:0045596: negative regulation of cell differentiation
R-MMU-2029481: FCGR activation
GO:0070838: divalent metal ion transport
GO:0071887: leukocyte apoptotic process
GO:0034393: positive regulation of smooth muscle cell apoptotic process
GO:0051482: positive regulation of cytosolic calcium ion concentration involved in phospholipase C

Fig. 3. Effect of dietary fiber intake in patients and in preclinical models of melanoma immunotherapy. (**A**) Kaplan-Meier plot comparing progression-free survival intervals by dietary fiber intake among patients who received ICB (n = 128; P = 0.047 by log-rank test). (**B**) Kaplan-Meier plot comparing progression-free survival intervals by combined dietary fiber and probiotic status among patients who received ICB (n = 123; overall P across four groups = 0.11; P for sufficient dietary fiber intake + no probiotics use versus else = 0.015; both by log-rank test). (**C**) Experimental design of studies in C57BL/6 SPF mice that received either a high-fiber or low-fiber diet at inoculation of M3 (HCmel1274) melanoma cells (1×10^6 tumor cells) and were then treated with anti–PD-1 or isotype control. Time is in days relative to tumor injection. (**D**) M3 melanoma growth kinetics of control (high-fiber) diet (circles) or low-fiber (fiber-free) diet (squares) treated four times with intraperitoneal injection of anti–PD-1 antibody (dark green) or of isotype control (Iso Ctrl) (light green). Data are means ± SEM of tumor volume from one representative experiment (n = 5 per group). All P values are from a likelihood ratio test in a linear mixed model (isotype control and high fiber, P = 0.69; anti–PD-1 and high fiber, P = 0.02; anti–PD-1 and low fiber versus isotype control and low fiber, P = 0.08). *P < 0.05. (**E**) t-UMAP plot comparing the gut microbiome (via shotgun metagenomic sequencing of fecal samples) of mice by treatment and diet group from two experiments (n = 4 to 5 per group) using Bray-Curtis distances (PERMANOVA P < 0.0001) at experimental day 16. (**F**) Heatmap of gene expression of flow-sorted $CD45^+$ tumor-infiltrating immunocytes in mice fed high- versus low-fiber diets and treated with anti–PD-1 or isotype control. (**G**) Gene set enrichment analysis depicting pathways enriched in high-fiber diet mice treated with anti–PD-1 versus isotype control which were not differentially expressed by treatment in low-fiber diet mice.

of fiber on antitumor immunity by affecting pathways of T cell activation as well as the accumulation of T cells in the tumor, including inducible T cell co-stimulator (ICOS)–expressing $CD8^+$ and $CD4^+$ T cells (fig. S17, C to H).

Together, these data have important implications. We show that dietary fiber and probiotic use, factors known to affect the gut microbiome, are associated with differential outcomes to ICB. Although causality cannot be addressed from the observational human cohort, where unmeasured confounders may exist, our preclinical models support the hypothesis that dietary fiber and probiotics modulate the microbiome and that antitumor immunity is impaired in mice receiving a low-fiber diet and in those receiving probiotics—with suppression of intratumoral IFN-γ T cell responses in both cases.

Numerous challenges exist to decipher how best to leverage the microbiome to optimize patient outcomes, starting with what to target—selected features or community function—and whether this can be safely achieved through supplementation or more comprehensive dietary approaches. Several prior studies have shown that controlled increases in dietary fiber intake can modulate the gut microbiome but also that interindividual variation in the gut microbiome drives differential effects of specific fibers (and prebiotics) on host metabolism (*36*–*40*). Ongoing dietary intervention studies in the setting of ICB are critical for establishing whether a targeted and achievable diet change at the initiation of ICB can safely and effectively improve outcomes (NCT04645680). Although our findings suggest that undirected use of commercially available probiotics may be harmful in the setting of ICB, further study of rationally designed and targeted probiotics or bacterial consortia is warranted on the basis of promising early data of this approach (*22*–*24*).

Some analyses in the current cohort were not adequately powered to assess the full effect of these factors, and further validation is needed in independent cohorts with more in-depth and detailed assessment of dietary intake and the use of specific probiotic supplements, along with further mechanistic studies in preclinical models. Nonetheless, these notable (and perhaps unexpected) findings from studies in this observational patient cohort are corroborated by parallel studies in preclinical models with preliminary mechanistic insights. In light of these collective results, dietary habits and probiotic supplement use should be considered in patients receiving ICB and in efforts to modulate the gut microbiota. These factors should be more thoughtfully evaluated in strategies to improve cancer outcomes.

REFERENCES AND NOTES

1. A. Ribas, J. D. Wolchok, *Science* **359**, 1350–1355 (2018).
2. B. Routy *et al*., *Science* **359**, 91–97 (2018).
3. V. Matson *et al*., *Science* **359**, 104–108 (2018).
4. V. Gopalakrishnan *et al*., *Science* **359**, 97–103 (2018).
5. A. E. Frankel *et al*., *Neoplasia* **19**, 848–855 (2017).
6. B. A. Peters *et al*., *Genome Med.* **11**, 61 (2019).
7. M. C. Andrews *et al*., *Nat. Med.* **27**, 1432–1441 (2021).
8. G. D. Wu *et al*., *Science* **334**, 105–108 (2011).
9. D. McDonald *et al*., *mSystems* **3**, e00031-18 (2018).
10. L. Maier *et al*., *Nature* **555**, 623–628 (2018).
11. J. Suez *et al*., *Cell* **174**, 1406–1423.e16 (2018).
12. B. P. Willing, S. L. Russell, B. B. Finlay, *Nat. Rev. Microbiol.* **9**, 233–243 (2011).
13. F. Asnicar *et al*., *Nat. Med.* **27**, 321–332 (2021).
14. D. Rothschild *et al*., *Nature* **555**, 210–215 (2018).
15. E. A. Eisenhauer *et al*., *Eur. J. Cancer* **45**, 228–247 (2009).
16. F. E. Thompson, D. Midthune, L. Kahle, K. W. Dodd, *J. Nutr.* **147**, 1226–1233 (2017).
17. National Cancer Institute, Dietary Screener Questionnaire (DSQ) in the NHANES 2009-10: Data processing and scoring procedures (2019); https://epi.grants.cancer.gov/nhanes/dietscreen/scoring.
18. R. Z. Gharaibeh, C. Jobin, *Gut* **68**, 385–388 (2019).
19. E. N. Baruch *et al*., *Science* **371**, 602–609 (2021).
20. D. Davar *et al*., *Science* **371**, 595–602 (2021).
21. J. C. Arthur *et al*., *Sci. Rep.* **3**, 2868 (2013).
22. Y. Tomita *et al*., *Cancer Immunol. Res.* **8**, 1236–1242 (2020).
23. A. Sivan *et al*., *Science* **350**, 1084–1089 (2015).
24. T. Tanoue *et al*., *Nature* **565**, 600–605 (2019).
25. H. J. Flint, K. P. Scott, S. H. Duncan, P. Louis, E. Forano, *Gut Microbes* **3**, 289–306 (2012).
26. G. den Besten *et al*., *J. Lipid Res.* **54**, 2325–2340 (2013).
27. S. A. Poeker *et al*., *Sci. Rep.* **8**, 4318 (2018).
28. D. So *et al*., *Am. J. Clin. Nutr.* **107**, 965–983 (2018).
29. M. S. Desai *et al*., *Cell* **167**, 1339–1353.e21 (2016).
30. J. L. McQuade *et al*., *Lancet Oncol.* **19**, 310–322 (2018).
31. Y. An *et al*., *J. Transl. Med.* **18**, 235 (2020).
32. D. J. Pinato *et al*., *JAMA Oncol.* **5**, 1774–1778 (2019).
33. D. M. Klurfeld *et al*., *Am. J. Physiol. Endocrinol. Metab.* **315**, E1087–E1097 (2018).
34. K. Makki, E. C. Deehan, J. Walter, F. Bäckhed, *Cell Host Microbe* **23**, 705–715 (2018).
35. E. Pérez-Guijarro *et al*., *Nat. Med.* **26**, 781–791 (2020).
36. L. Michalak *et al*., *Nat. Commun.* **11**, 5773 (2020).
37. A. Oliver *et al*., *mSystems* **6**, e00115-21 (2021).
38. A. Leshem, E. Segal, E. Elinav, *mSystems* **5**, e00665-20 (2020).
39. S. M. Murga-Garrido *et al*., *Microbiome* **9**, 117 (2021).
40. C. R. Daniel, J. L. McQuade, *Trends Cancer* **5**, 521–524 (2019).

ACKNOWLEDGMENTS

C.N.S. acknowledges the Parker Institute for Cancer Immunotherapy for funding time devoted to continued analysis of this project. She also acknowledges M. D. Swartz, L. B. Piller, and X. Du at UTSPH for their participation on her thesis committee. J.L.M. acknowledges the Transdisciplinary Research in Energetics and Cancer Research Training Workshop R25CA203650. J.L.M. and C.R.D. acknowledge the MDACC Center for Energy Balance in Cancer Prevention and Survivorship. C.R.D. acknowledges the MDACC Bionutrition Core. We thank our collaborators and personnel at the LICI, NCI Microbiome and Genetics Core Facility, and NCI Mouse Gnotobiotic Core Facility, The Alkek Center for Metagenomics and Microbiome Research (CMMR) at Baylor College of Medicine (including S. J. Javornik Cregeen for her role in microbiome data processing), and CosmosID for their high quality and timely microbiome data generation services. We also wish to acknowledge MD Anderson's Program for Innovative Microbiome and Translational Research (PRIME-TR) for supporting the analysis and interpretation of the microbiome results presented herein (J.A.W. and N.J.A. are the program director and executive scientific director for PRIME-TR, respectively). Most importantly, the study team wishes to thank all patients who contributed their time, samples, and data to this research.

Funding: This study received support from National Institute of Health grant 1R01 CA219896-01A1 (J.A.W.); US-Israel Binational Science Foundation grant 201332 (J.A.W.); the Melanoma Research Alliance grant 4022024 (J.A.W.); American Association for Cancer Research Stand Up to Cancer grant SU2C-AACR-IRG-19-17 (J.A.W.); the Andrew Sabin Family Fellows Program (J.A.W. and C.R.D.); MD Anderson Cancer Center's Melanoma Moon Shots Program (J.A.W., J.L.M., C.R.D., M.A.D., H.A.T., J.E.G., and E.M.B.); and the Melanoma Research Alliance grant 564449 (L.C., J.A.W., and J.L.M.). The authors additionally received support from Department of Defense grant W81XWH 16 1 0121 (J.A.W.); the MD Anderson Cancer Center Multidisciplinary Research Program grant (J.A.W.); the Parker Institute for Cancer Immunotherapy at MD Anderson Cancer Center (J.A.W., H.A.T., P.S., and J.P.A.); American Society of Clinical Oncology and Conquer Cancer Foundation Career Development award AWD0000627 (J.L.M.); the Elkins Foundation (J.L.M.); the Seerave Foundation (J.L.M.); Rising Tide Foundation grant AWD00004505 (J.L.M.); the Mark Foundation grant AWD00004538 (J.L.M.); the Longenbaugh-Torian Fund (J.L.M.); the MD Anderson Cancer Center SPORE in Melanoma P50CA221703 (M.A.D., H.A.T., I.C.G., J.A.W., S.P.P., J.E.L., J.E.G., A.J.L., and J.L.M.); the MD Anderson Physician Scientist Program (J.L.M.); the Cancer Research Institute Irvington Fellowship Program (M.V.); Cancer Prevention and Research Institute of Texas Research Training Program RP170067 (A.P.C., J.T., and J.Z.); Cancer Prevention and Research Institute of Texas Research Training Program RP210028 (L.M.K.); the US Department of State, Bureau of Educational and Cultural Affairs (A.P.C.); the Fulbright Franco–Américaine Commission (A.P.C.); Cancer Prevention and Research Institute of Texas Research

Training Program RP160097 (X.Z.); National Health and Medical Research Council of Australia CJ Martin Early Career Fellowship grant 1148680 (M.C.A.); National Institutes of Health T32 CA 009599 (M.G.W. and B.A.H.); the MD Anderson NCI Cancer Center Support grant P30 CA016672 (M.G.W., B.A.H., C.R.D., K.C.-D., and C.B.P.); National Institute of Health grant 1F32CA260769-01 (G.Mo.); the Charles A. King Trust Postdoctoral Research Fellowship (Y.Y. and C.H.); the John M. Skibber Endowed Professorship (J.E.G.); the Michael and Patricia Melanoma Research Endowment (J.E.G.); National Institute of Health grant R0AI143886 (J.H.); Columbia University Health Sciences NCI Cancer Center Support grant P30 CA013696 (J.H.); National Institute of Health grant R01HL124112 (R.R.J.); Cancer Prevention and Research Institute of Texas grant RR160089 (R.R.J.); National Institute of Health grant R01AI109294 (S.S.W.); National Institute of Health grant R01AI133822 (S.S.W.); the Richard E. Haynes Distinguished Professor in Clinical Cancer Prevention (L.C.); American Cancer Society grant RSG-17-049-01-NEC (C.R.D.); the National Institute of Health Intramural Research Program (E.P.-G., C.-P.D., and G.Me.); FLEX Synergy Award from the National Cancer Institute Center for Cancer Research (E.P.-G., C.-P.D., and G.Me.); the National Cancer Institute Intramural Research Program (J.A.M., M.V., J.H.B., R.R.R., and G.T.); the Dr. Miriam and Sheldon G. Adelson Medical Research Foundation (M.A.D.); American Cancer Society/ Melanoma Research Alliance grant 134148-MRAT-19-168-01 (M.A.D.); the AIM at Melanoma Foundation (M.A.D.); the National Institute of Health/National Cancer Institute grant 1 P50 CA221703-02 (M.A.D.); the National Institute of Health/National Cancer Institute grant 1U54CA224070-03 (M.A.D.); Cancer Fighters of Houston (M.A.D.); and the Anne and John Mendelsohn Chair for Cancer Research (M.A.D.). **Author contributions:** Project conceptualization: C.N.S., J.L.M., C.H., M.V., N.J.A., G.T., L.C., W.S.G., C.R.D., S.S.W., J.A.W., M.I.R., Y.S., and M.A.W.K. Methodology: C.N.S., J.L.M., N.J.A., V.G., C.H., M.V., J.C.M., G.T., L.C., C.R.D., J.A.W., K.L.H., J.A.M., C.B.P., J.H., A.J.L., and M.A.W.K. Software: C.N.S., V.G., M.A.W.K., J.Z., X.Z., K.L.H., J.F.P., J.R., N.J.A., M.C.W., L.E.H., Y.Y., C.H., J.A.M., J.H.B., R.R.R., and A.M. Validation: M.V., J.A.M., C.N.S., V.G., M.A.W.K., J.Z., X.Z., A.P.C., S.S.W., Y.Z., Y.B.M., S.J., L.N., S.J., and H.A.T. Formal analysis: C.N.S., V.G., A.P.C., M.A.W.K., J.Z., X.Z., K.L.H., J.R., N.J.A., Y.Y., C.H., J.A.M., R.R.R., A.M., C.B.P., J.H., C.R.D., L.W., L.N., and N.-A.A.S.A. Investigation: C.N.S., J.L.M., B.A.H., A.P.C., P.-O.G., C.W.H., K.W., S.S.W., Y.Z., Y.B.M., M.V., Y.S.K., E.P.-G., I.C.G., R.N.A., H.A.T., A.D., M.K.W., S.P.P., S.E.W., M.A.D., M.I.R., J.E.G., J.E.L., P.H., V.J., K.C.N., A.J.L., M.T.T., L.M.K., A.L.H., M.A.J., E.N.B., B.P., B.S.L., K.C.-D., T.C., S.S., N.-A.A.S.A., J.T., V.J., C.M.Z., M.A.W.K., and J.A.W. Resources: J.A.W., J.P.A., P.S., S.S.W., R.R.J., C.H., G.T., J.F.P., P.A.F., G.Me., E.P.-G., C.-P.D., A.J.L., B.S.L., and H.A.T. Data curation: M.W., N.J.A., J.A.M., R.R.R., A.M., C.N.S., J.L.M., C.R.D., E.S., E.P., L.E.H., J.C.M., T.R., E.N.B., R.A.S.-S., J.A., C.M.Z., L.W., M.C.W., M.C.A., N.-A.A.S.A., and B.S.L. Writing - original draft: C.N.S., J.L.M., V.G., J.A.M., M.V., A.P.C., M.A.W.K., N.J.A., L.N., S.S.W., L.C., G.T., C.R.D., J.A.W., J.H., J.P.A., and P.S. Writing - review and editing: C.N.S., J.L.M., V.G., J.A.M., M.V., A.P.C., M.A.W.K., X.Z., M.G.W., C.B.P., M.C.W., G.Me., T.R., J.H.B., B.A.H., M.C.A., R.R.R., A.M., Y.S.K., J.R., K.L.H., J.Z., Y.Z., Y.B.M., L.M.K., S.J., C.W.H., K.W., P.-O.G., A.L.H., M.A.J., E.N.B., E.P.-G., C.-P.D., G.Mo., B.P., B.S.L., R.A.S.-S., R.A., J.A., C.M.Z., E.P., E.S., J.S., L.E.H., E.M.B., L.W., M.D., K.C.-D., S.S., T.C., N.-A.A.S.A., S.D., J.T., J.C.M., I.C.G., R.N.A., H.A.T., A.D., M.K.W., S.P.P., S.E.W., M.A.D., M.I.R., J.E.G., J.E.L., P.H., V.J., Y.S., R.S., N.J.A., K.C.N., L.N., J.F.P., P.A.F., A.J.L., J.H., R.R.J., R.R.R., Y.Y., W.S.G., C.H., P.S., S.S.W., J.P.A., L.C., G.T., C.R.D., and J.A.W. Visualization: C.N.S., J.L.M., V.G., J.A.M., M.V., A.P.C., M.A.W.K., X.Z., N.J.A., R.R.R., A.M., C.B.P., J.H., C.R.D., J.A.W., G.T., Y.Y., C.H., K.C.-D., and M.I.R. Supervision: C.N.S., J.L.M., V.G., L.C., M.V., G.T., C.H., C.R.D., J.A.W., and P.A.F. Project administration: C.N.S., J.L.M., M.G.W., S.D., E.M.B., J.S., L.E.H., C.R.D., J.A.W., M.V., G.T., and G.Me. Funding acquisition: J.L.M., C.R.D., L.C., J.A.W., E.M.B., G.T., M.A.D., C.H., and Y.S. **Competing interests:** J.A.W., V.G., and M.C.A. are inventors on patent WO2020106983A1 submitted by the Board of Regents, The University of Texas System, and Institut Gustav Roussy that covers methods and compositions for treating cancer and for predicting a subject's response to combination checkpoint inhibitor therapy. J.A.W. and V.G. are inventors and C.N.S. is a collaborator on a US patent application (PCT/US17/53.717) submitted by the University of Texas MD Anderson Cancer Center that covers methods to enhance ICB responses by modulating the microbiome. J.A.W. and R.R.J. are inventors on patent WO2020150429A1 submitted by the Board of Regents, The University of Texas System, that covers methods and compositions for treating immune checkpoint inhibitor (ICI)–associated colitis in a subject through the administration of fecal matter from a healthy donor to the subject. R.R.J. is an inventor on patent WO2016086161A1 submitted by the Memorial Sloan Kettering Cancer Center that covers compositions and methods for increasing the abundance of commensal bacteria belonging to the order Clostridiales that are associated with reduced lethal graft-versus-host disease and improved overall survival after bone marrow or hematopoietic stem cell transplant. R.R.J. is an inventor on patent WO2017041039A1 submitted by Memorial Sloan Kettering Cancer Center that covers methods and compositions for reducing the risk of cancer relapse in a subject who has received cancer treatment. J.A.W. reports compensation for speaker's bureau and honoraria from Imedex, Dava Oncology, Omniprex, Illumina, Gilead, PeerView, Physician Education Resource, MedImmune, and Bristol-Myers Squibb and serves as a consultant and/or advisory board member for Roche/ Genentech, Novartis, AstraZeneca, GlaxoSmithKline, Bristol-Myers Squibb, Merck, Biothera Pharmaceuticals, Microbiome DX, and Micronoma. J.A.W. also receives research support from GlaxoSmithKline, Roche/Genentech, Bristol-Myers Squibb, and Novartis. V.G. reports honoraria from ExpertConnect and Kansas Society of Clinical Oncology. J.L.M. reports advisory board participation and honoraria from Bristol-Myers Squibb, Merck, and Roche/Genentech. M.C.A. reports advisory board participation, honoraria, and research funding to their institution from MSD Australia, outside the submitted work, and contract research for BMS Australia, outside the current work. A.P.C. is an employee and equity holder at Immunai. A.P.C. serves as an advisory member and holds equity in Vastbiome. G.Mo. is a coinventor on US patents (PCT/US2019/022194, PCT/US2020/029556, and PCT/US2020/ 046050) not related to the content of this paper. M.I.R. reports paid consultant position for AMGEN and paid consultant and advisory board membership for MERCK. I.C.G. reports research support from Bristol-Myers Squibb and Merck. I.C.G. serves as a consultant for Bristol-Myers Squibb, Array, and Novartis. R.N.A. receives research funding from Merck, Bristol-Myers Squibb, Genentech, Novartis, and Iovance. J.E.G. is a consultant and/or is on the advisory board for Merck, Regeneron, Syndex, Novartis, and Bristol-Myers Squibb, unrelated to the content of this work. R.R.J. is an advisor and holds equity in Seres Therapeutics and Kaleido Biosciences; serves on the advisory board of MaaT Pharma, LISCure Biosciences, and Prolacta Biosciences; and consults for Davolterra, Merck, Microbiome DX, and Karius. C.H. is on the scientific advisory board for Seres Therapeutics and Empress Therapeutics. M.K.W. is on the advisory boards of Merck, Pfizer, Bristol Myers Squibb, Regeneron, EMD-Serono, ExiCure, Castle Biosciences, and Adagene. H.A.T. is a consultant for BMS, Merck, Novartis, Genentech, Eisai, Iovance, Karyopharm, and Pfizer and reports research funding to institution from BMS, Merck, Novartis, Genentech, GSK, and Dragonfly. A.D. serves as a consultant for Nektar, MultiVir, Idera, Array, and Bristol-Myers Squibb. P.S. reports consulting or stock ownership or advisory board for Achelois, Adaptive Biotechnologies, Affini-T, Apricity, BioAtla, BioNTech, Candel Therapeutics, Catalio, Codiak, Constellation, Dragonfly, Earli, Enable Medicine, Glympse, Hummingbird, ImaginAb, Infinity Pharma, Jounce, JSL Health, Lava Therapeutics, Lytix, Marker, Oncolytics, PBM Capital, Phenomic AI, Polaris Pharma, Sporos, Time Bioventures, Trained Therapeutix, Two Bear Capital, and Venn Biosciences. J.P.A. reports consulting, stock ownership, or advisory board membership for Achelois, Adaptive Biotechnologies, Apricity, BioAtla, BioNTech, Candel Therapeutics, Codiak, Dragonfly, Earli, Enable Medicine, Hummingbird, ImaginAb, Jounce, Lava Therapeutics, Lytix, Marker, PBM Capital, Phenomic AI, Polaris Pharma, Time Bioventures, Trained Therapeutix, Two Bear Capital, and Venn Biosciences. M.A.D. has been a consultant to Roche/ Genentech, Array, Pfizer, Novartis, BMS, GSK, Sanofi-Aventis, Vaccinex, Apexigen, Eisai, and ABM Therapeutics, and he has been the principal investigator of research grants to MD Anderson from Roche/Genentech, GSK, Sanofi-Aventis, Merck, Myriad, and Oncothyreon. P.H. is on the scientific advisory board for Dragonfly and Immatics. M.I.R. is on the melanoma advisory board for Merck and is a paid consultant for AMGEN and Merck. S.P.P. reports institutional clinical trial support from NCI, Merck, and Bristol Myers Squibb during the conduct of the study; institutional clinical trial support from Reata Pharmaceuticals, Novartis, Deciphera, Provectus Biopharmaceuticals, Foghorn Therapeutics, TriSalus Life Sciences, and Seattle Genetics; advisory board honoraria from Castle Biosciences and TriSalus Life Sciences; honoraria as Peer Discussion Group Leader for Merck; and honoraria for service as Chair of International Data Monitoring Committee for Immunocore, outside the submitted work. **Data and materials availability:** Raw sequencing data and all relevant human data necessary for reproducing results are available in the NCBI Sequence Read Archive under BioProject ID PRJNA770295. All analyzed sequencing data are available in the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.aaz7015
Materials and Methods
Figs. S1 to S18
Tables S1 to S14
References (*41*–*74*)
MDAR Reproducibility Checklist

1 October 2019; resubmitted 17 May 2020
Accepted 24 November 2021
10.1126/science.aaz7015

WAYNE STATE UNIVERSITY

IBio

METABOLIC DISORDERS AND CARDIOVASCULAR HEALTH

As part of a broad institutional initiative in integrative biomedical sciences (IBio), Wayne State University is recruiting for **faculty positions** (tenured or tenure-track, open rank) in the integrative biology of cardiometabolic health and disease. The recruitment seeks to complement existing IBio strengths in cell biology, lipid metabolism, and cardiometabolic physiology. Of particular interest are investigators addressing molecular and cellular processes that impact energy metabolism, metabolic signaling, and cardiovascular function.

Successful applicants will benefit from a highly collaborative research environment working with colleagues throughout the University. Candidates must have a Ph.D., M.D., or equivalent degree in biomedical sciences relevant to metabolic diseases, with evidence of peer recognition in the field, a commitment to excellence in research education and training, and the ability to engage with broader integrative science themes for the purpose of achieving transformative and translational research gains. Applicants are expected to have already established extramural research funding or be on a clear path to secure and sustain extramural funding in support of their research programs. Faculty recruits will integrate with departments and colleges or schools consistent with their areas of expertise and interests and engage in all aspects of our academic mission, including research, education and service.

Competitive recruitment packages are available with salary and faculty rank based on qualifications. Applicants are encouraged to apply to posting **#046049** through the Online Hiring System **https://jobs.wayne.edu/applicants/jsp/shared/Welcome_css.jsp**

Applications will be accepted until positions are filled, but for full consideration, application materials should be submitted by **February 15, 2022**. Applications should include a curriculum vitae and a brief cover letter to the Assistant Vice President for Integrative Biosciences indicating the applicant's potential for synergy in the targeted thematic areas.

Wayne State University is a premier, public, urban research university located in the heart of Detroit where students from all backgrounds are offered a rich, high quality education. Our deep-rooted commitment to excellence, collaboration, integrity, diversity and inclusion creates exceptional educational opportunities preparing students for success in a diverse, global society. Wayne State University encourages applications from women, people of color and other underrepresented people. WSU is an affirmative action/equal opportunity employer.

Founded in 1868, Wayne State University offers a range of academic programs through 13 schools and colleges to nearly 28,000 students. The campus in Midtown Detroit comprises 100 buildings over 200 acres including the School of Medicine, the College of Engineering, the Eugene Applebaum College of Pharmacy and Health Sciences and the College of Nursing. The university is home to the Perinatology Research Branch of the National Institutes of Health, the Karmanos Cancer Center, a National Cancer Institute-designated comprehensive cancer center, and a National Institute of Environmental Health Sciences Core Center - Center for Urban Responses to Environmental Stressors (CURES).

ScienceCareers

FROM THE JOURNAL SCIENCE AAAS

Confused about your next career move?

Download Free Career Advice Booklets!

ScienceCareers.org/booklets



THE GEORGE WASHINGTON UNIVERSITY
WASHINGTON, DC

INTRODUCTORY BIOLOGY TEACHING FACULTY

Position Description Summary
The Department of Biological Sciences at The George Washington University (GWU) invites applicants for a teaching **faculty position in Introductory Biology**. This is a full-time, renewable contract position beginning Fall 2022. The successful candidate will be actively involved in the design, delivery, and supervision of undergraduate laboratories associated with our Introductory Biology sequence for majors. The candidate will teach a large lecture section in either the majors or non-majors Introductory Biology course each semester, as needed. The candidate will manage and coordinate labs and course topics while collaborating with our teaching faculty, laboratory managers, graduate teaching assistants, and undergraduate teaching assistants to ensure a high-quality laboratory experience that prepares our majors for advanced coursework in Biology. We particularly welcome candidates from groups that are historically underrepresented in biology and/or have demonstrated leadership toward building an equitable and inclusive scholarly environment.

To be considered, please complete an online faculty application (**http://www.gwu.jobs/postings/88600**). Review of applications will begin on **January 10th, 2022**, and continue until the position is filled, pending final budgetary approval. Employment offers are contingent on the satisfactory outcome of a standard background screening.

The university is an Equal Employment Opportunity/Affirmative Action employer that does not unlawfully discriminate in any of its programs or activities on the basis of race, color, religion, sex, national origin, age, disability, veteran status, sexual orientation, gender identity or expression, or on any other basis prohibited by applicable law.

By **Jennifer Lanni**

A lesson in failure

With class about to start, I print 14 Western blot images for my students to discuss. The 3-hour lab is supposed to be the culmination of a weekslong research project in my undergraduate biology course, the day my students determine whether their experimental results support their carefully crafted hypotheses. But the images are all the same—and all full of nothing but background bands. My students are about to have a hard lesson in scientific failure and how to be resilient in the face of it. It's a lesson I wish I'd learned before starting grad school.

As a Ph.D. student, I went through many frustrating cycles nurturing hypotheses that would later wither because of technical failures or ambiguous experimental results. When I graduated, I thought I would never do bench research again. I wasn't willing to sacrifice so many hours away from my young children with so little to show for it. Instead, I embraced my love of teaching by working as a lecturer.

A decade later and with my kids in school, my scientific curiosity came out of hibernation and I restarted my research career. After a stint as a postdoc, I secured a position at a liberal arts college, where I established my own small lab. In my undergraduate classes, I asked my students to complete lab experiments that were virtually guaranteed to yield interpretable data. But after a few years, I grew uncomfortable with the gap between those picture-perfect experiments and my own research projects. Yes, my students left lab sessions with results, feeling their time had been well-spent. But I worried I was deceiving them about the actual experience of practicing science, which rarely produces data on the first try.



"Research is messy. ... Trying to protect students from that reality does them a disservice."

I decided to develop a new course that would give our students experience performing real experiments, ones that had the potential to fail. Using my own research interests as a framework, I gave the students a collection of papers to read. During group brainstorming sessions, they identified new questions that arose from what had already been done and collectively came up with their own hypotheses. After spending time learning lab methods required to test their hypotheses, they got to work performing their first experiment.

On the day of data analysis, I handed them the Western blot printouts and asked them to look over the images and discuss their findings. Most assumed their blots were correct—that the background bands they saw represented the proteins they had hoped to detect—and jumped immediately to interpreting the data. But I refused to let the students move on.

After a solid hour of struggle and some leading questions on my part, one student finally spoke up. "It doesn't make sense. The bands look the same size, but the proteins should be different sizes." Hallelujah! A student had stepped back from seeing what they expected to see and described what the data actually showed. Their breakthrough helped their classmates start to look at the results with more objective eyes. Within minutes, they were overflowing with questions and ideas about what could have gone wrong. We spent the next 2 hours covering the chalkboard with plans to troubleshoot the experimental procedures. My students were thinking like scientists—a development no amount of advance planning could have created.

Afterward, I reflected on how we train future scientists. Should we talk more openly with students about failure? When I quietly left research, frustrated at what felt like my lack of accomplishment, was this a typical experience? How often do we inadvertently discourage students from persisting in science, simply by omitting honest descriptions of the failure inherent to the research process? Research is messy and full of failed attempts. Trying to protect students from that reality does them a disservice.

My class never did generate data to test their hypotheses. Instead, we finished the semester reading about and discussing scientific failure. I hope the handful of students who go on to graduate school will learn from the experience and bring an awareness with them that success doesn't always come easy. I trust this knowledge will help my other students as well, wherever they go. ■

Jennifer Lanni is an associate professor at Wheaton College in Massachusetts. Send your career story to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER



Connectivity Reshaping the Future of Health Care

Mindray, a leading global medical device and solution provider headquartered in Shenzhen, China, focuses on advancing and sharing health technology to bring better care to more people across the globe. Mindray's primary operations cover patient monitoring and life support, in-vitro diagnostics, and medical imaging systems. Through its innovative M-Connect IT solution and MiCo+, Mindray, working with Intel, is committed to helping healthcare providers boost efficiency and optimize management.



At the First Affiliated Hospital of the Zhejiang University School of Medicine, the M-Connect IT solution for the ICU unlocks the potential of patient monitoring by creating a universal, central monitoring platform. It enables the hospital to achieve seamless device integration, greater data visibility, streamlined workflows, and efficient clinical decision-making processes.

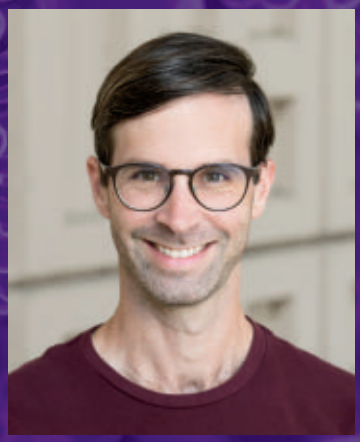
With the technical support of Mindray's remote imaging platform MiCo+, Shenzhen People's Hospital and Shenzhen Shekou People's Hospital fully meet the demands of ultrasonic remote quality control, teaching and training, and remote consultation.



https://www.mindray.com/en/

TRILLIONS OF MICROBES ONE ESSAY

The NOSTER *Science* Microbiome Prize is an international prize that rewards innovative research by investigators, under the age of 35, who are working on the functional attributes of the microbiota. The research can include any organism that has potential to contribute to our understanding of human or veterinary health and disease, or to guide therapeutic interventions. The winner and finalists will be chosen by a committee of independent scientist, chaired by a senior editor of *Science*. The top prize includes a complimentary membership to AAAS, an online subscription to *Science*, and $25,000 (USD). Submit your research essay today.



Eran Blacher, Ph.D.
2021 Winner

NOSTER | Science
MICROBIOME
PRIZE

Apply by 1/24/22 at www.science.org/noster

Sponsored by Noster, Inc