Tougher tests to make AI smarter and less biased p. 570

Low genetic diversity not a threat to vaquitas pp. 574 & 635

Subglacial sediments hold large volumes of water pp. 577 & 640

Science

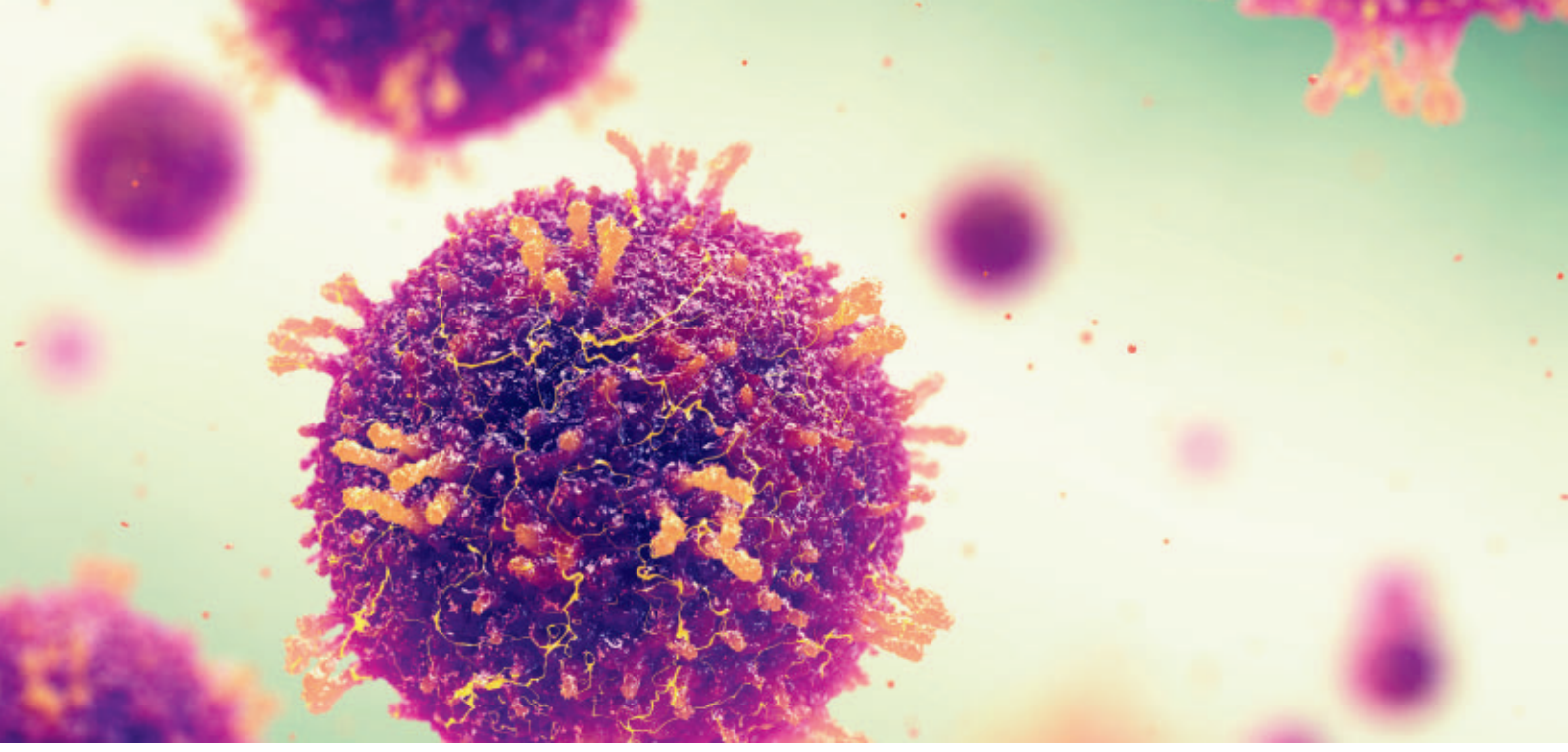
$15
6 MAY 2022
science.org

AAAS



SUNSTRUCK

Sunscreen turns toxic in cells of reef inhabitants pp. 578 & 644



It's time to think differently.

Find out how NEB can support your infectious disease research and development.

Gaining a better understanding of infectious diseases, including their characterization, evolution and transmission, continues to be a priority, both from an R&D standpoint and as a public health issue. The COVID-19 pandemic has demonstrated the need for a wide range of tools to research infectious diseases, and has highlighted the importance of speed and the ability to pivot as new problems arise. This has emphasized the need for innovation and thinking differently about where to access those critical materials, including genomics reagents.

Many scientists know NEB as a trusted reagent supplier to the life science community, but what you may not know is that we also offer a portfolio of products that can be used in infectious disease research, development of diagnostics and therapies, and in epidemiological studies and disease surveillance. In fact, many of our products have supported the development of COVID-19 diagnostics and vaccines, and can also be utilized with other infectious diseases, such as influenza and malaria.

- Benefit from almost 50 years of experience in molecular biology & enzymology
- Partner with our OEM & Customized Solutions team to find the best solution for your needs
- Take advantage of our expanded manufacturing capabilities
- Access product formats, such as GMP-grade*, lyophilized, lyo-ready and glycerol-free
- Be confident in your product performance with our expanded quality and regulatory systems

Ready to get started? Learn more at
www.neb.com/InfectiousDiseases

"GMP-grade" is a branding term NEB uses to describe reagents manufactured at our Rowley, MA facility, where we utilize procedures and process controls to manufacture reagents in compliance with ISO 9001 and ISO 13485 quality management system standards. NEB does not manufacture or sell products known as Active Pharmaceutical Ingredients (APIs), nor do we manufacture products in compliance with all of the Current Good Manufacturing Practice regulations.

One or more of these products are covered by patents, trademarks and/or copyrights owned or controlled by New England Biolabs, Inc. For more information, please email us at busdev@neb.com. The use of these products may require you to obtain additional third party intellectual property rights for certain applications.
© Copyright 2022, New England Biolabs, Inc.; all rights reserved.

NEW ENGLAND BioLabs® Inc.

be INSPIRED
drive DISCOVERY
stay GENUINE

CONTENTS



570

6 MAY 2022 • VOLUME 376 • ISSUE 6593

NEWS

IN BRIEF

558 News at a glance

IN DEPTH

562 Bids for Anthropocene's 'golden spike' emerge
Sites compete to mark global changes of the 1950s and define new geological age *By P. Voosen*
PODCAST

563 Census aims for better U.S. statistical portrait
Agency wants to retool its surveys and decennial census to improve efficiency and generate better data *By J. Mervis*

565 Doubt cast on inflammation's stop signals
Critics challenge data underpinning "resolution immunology," triggering university probes *By G. Sinha*

567 Germany weighs whether culling excess lab animals is a crime
As prosecutors evaluate complaints from animal rights groups, labs try to reduce surplus *By H. Feldwisch-Drentrup*

568 Balloon detects first signs of 'sound tunnel' in the sky
Atmospheric analog to ocean's acoustic channel could be used to monitor eruptions and bombs *By Z. Savitsky*

FEATURES

570 Taught to the test
AI software clears high hurdles on IQ tests but still makes dumb mistakes. Can better benchmarks help?
By M. Hutson

INSIGHTS

PERSPECTIVES

574 Using genomics to fight extinction
Quantifying fitness of wild organisms from genomic data alone is a challenging frontier
By C. E. Grueber and P. Sunnucks
REPORT p. 635

575 Giving entangled photons new colors
A fiber-based modulator shifts photon frequency while preserving quantum correlations *By A. V. Sokolov*
REPORT p. 621

577 Groundwater under Antarctica goes deep
A vast fossil seawater reservoir has been found beneath the West Antarctic Ice Sheet *By W. Chu*
REPORT p. 640

578 Sunscreens threaten coral survival
Oxybenzone-based sunscreens are increasing the mortality rates of stressed corals *By C. M. Hansel*
REPORT p. 644

579 Rapid response modeling of SARS-CoV-2 transmission
What can modelers learn from recent history to help prepare for the next pandemic?
By J. Zelner and M. Eisenberg
RESEARCH ARTICLE P. 596

581 Bacterial species rarely work together
Competition is prevalent and could be harnessed as an alternative to antibiotics
By J. D. Palmer and K. R. Foster

POLICY FORUM

583 Save the Mekong Delta from drowning
Policy must address drivers, not just symptoms, of subsidence
By G. M. Kondolf et al.

BOOKS *ET AL.*

586 The changing story of seafood
New technology, new markets, and better policy are improving fisheries and aquaculture
By O. P. Jensen

587 Hidden influences
Beliefs about the vital roles played by spirits and gods have long been underappreciated, argues an anthropologist *By J. Spencer*

LETTERS

588 Tree planting goals must account for wildfires
By A. B. Leverkus et al.

ILLUSTRATION: JASON SOLO/THE JACKY WINTER GROUP



THE ROAD TO SUCCESS IS PAVED WITH FAILURE.

UNTIL IT'S NOT.

Immune checkpoint inhibition, especially in combination with other therapeutic agents, has shown promise in the treatment of a variety of cancers. To successfully develop such therapies, researchers must have reliable tools to detect the checkpoint proteins and ligands in tissue, in addition to other therapeutic readouts. Shown here is paraffin-embedded human B-cell non-Hodgkin lymphoma stained for the immune checkpoint protein CTLA-4, using CTLA-4 (E2V1Z) Rabbit mAb #53560.

OUR PROCESS — YOUR SUCCESS.

When researchers expressed frustration that the existing CTLA-4 antibodies on the market weren't producing reliable IHC results, CST went to work and developed a high-performing option. As a company founded and led by active scientists, we understand the importance of having products you can count on. We validate every one of our thousands of antibodies and research products in-house—and in multiple applications where indicated. If any product fails to meet our stringent standards, we keep refining it until it does. Scientists all over the world trust CST® products to *work*—first time, every time. You can, too.

www.cellsignal.com

Cell Signaling TECHNOLOGY®

© 2022 Cell Signaling Technology, Inc. Cell Signaling Technology, CST, and XP are trademarks or registered trademarks of Cell Signaling Technology, Inc. 22-BCH-37353-MAY22

589 Adapt biodiversity targets to climate change
By A. V. Liz et al.

589 Outside the Tower: Air quality education in public schools
By H. A. Al-Abadleh et al.

590 Technical Comment abstracts

RESEARCH

IN BRIEF

591 From *Science* and other journals

REVIEW

594 Wetland ecology
Recovering wetland biogeomorphic feedbacks to restore the world's biotic carbon hotspots
R. J. M. Temmink et al.
REVIEW SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABN1479

RESEARCH ARTICLES

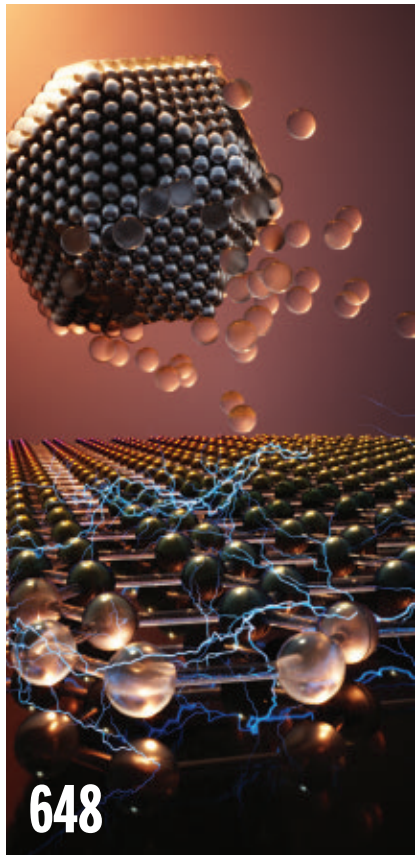
595 Neuroscience
The geometry of domain-general performance monitoring in the human medial frontal cortex
Z. Fu et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABM9922

596 Coronavirus
Increased risk of SARS-CoV-2 reinfection associated with emergence of Omicron in South Africa
J. R. C. Pulliam et al.
RESEARCH ARTICLE SUMMARY; FOR FULL TEXT: DOI.ORG/10.1126/SCIENCE.ABN4947
PERSPECTIVE p. 579

597 Climate economics
Land-use emissions embodied in international trade *C. Hong* et al.

603 Catalysis
The state of zinc in methanol synthesis over a Zn/ZnO/Cu(211) model catalyst
P. Amann et al.

609 Physiology
Z-DNA binding protein 1 promotes heatstroke-induced cell death
F. Yuan et al.



Ferroelectricity is a rare form of electricity that unexpectedly emerges at reduced dimensions on silicon in atomically thin zirconia.

REPORTS

615 Catalysis
Highly efficient catalytic production of oximes from ketones using in situ–generated H_2O_2
R. J. Lewis et al.

621 Quantum optics
Tunable and state-preserving frequency conversion of single photons in hydrogen *R. Tyumenev* et al.
PERSPECTIVE p. 575

624 Hydrogels
Cuticular pad–inspired selective frequency damper for nearly dynamic noise–free bioelectronics
B. Park et al.

630 Evolution
Idiosyncratic epistasis leads to global fitness–correlated trends
C. W. Bakerlee et al.

635 Conservation
The critically endangered vaquita is not doomed to extinction by inbreeding depression
J. A. Robinson et al.
PERSPECTIVE p. 574

640 Ice sheets
A dynamic saline groundwater system mapped beneath an Antarctic ice stream *C. D. Gustafson* et al.
PERSPECTIVE p. 577

644 Ecotoxicology
Conversion of oxybenzone sunscreen to phototoxic glucoside conjugates by sea anemones and corals
D. Vuckovic et al.
PERSPECTIVE p. 578; PODCAST

648 Ferroelectrics
Emergent ferroelectricity in subnanometer binary oxide films on silicon
S. S. Cheema et al.

653 Paleoecology
A stronger role for long-term moisture change than for CO_2 in determining tropical woody vegetation change
W. D. Gosling et al.

DEPARTMENTS

555 Editorial
Faculty must lead inclusion
By F. A. Hrabowski III

662 Working Life
What I promised my mother
By E. Virguez

ON THE COVER

Coral reefs, such as this site off the coast of Mayotte in the Indian Ocean, are under threat from many human activities. Biochemical experiments in a sea anemone model system now show that the sunscreen component oxybenzone is converted into toxic photosensitizers in reef-dwelling organisms. See pages 578 and 644. *Photo: Gabriel Barathieu/ Minden Pictures*



Science Staff 552
New Products 657
Science Careers 658

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 © 2022 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): $2212; 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.

ILLUSTRATION: UC BERKELEY IMAGE BY ELLA MARU STUDIO



Protecting the planet with biotechnology

The BioInnovation Institute Foundation (BII), Novozymes, and EvodiaBio turn molecular innovations into processes and products that make life on Earth more sustainable for the environment and healthier for people.

Between California and Hawaii, there's an island of plastic almost as big as Alaska. Plastic waste evident in this huge mass—known as the Great Pacific Garbage Patch—and atmospheric carbon dioxide (CO_2) at the highest levels in hundreds of thousands of years provide just two examples of the ongoing danger to Earth's sustainability. In Copenhagen, Denmark, scientists at BII use biotechnology to take on these environmental challenges and others.

"I think everyone knows that the planet's current health is not very strong," says BII CEO Jens Nielsen. He adds that in addition to the oceanic and climatic indicators, we are "overusing all the resources on the planet, and so we will have to find new solutions to deal with this."

In conjunction with its traditional focus on human health, BII is creating a new emphasis on planetary health. "It's very important that we make sure that we can all live healthy longer and that there's also a planet that is nice to live on," Nielsen says. As it turns out, these two objectives can often be addressed in parallel.

Farm to fork

Today's agricultural industry offers several opportunities where biotechnology can improve the health of the planet and people. According to the U.S. Department of Agriculture's Economic Research Service, more than 10% of the country's greenhouse gases come from agriculture—as one example, synthetic fertilizers release nitrous oxide. But that could change. "Imagine if we could add microbes to the soil that work as fertilizer?" Nielsen asks. "The microbes could directly take the nitrogen from the air and donate it to the plants."

Scientists at Novozymes in Copenhagen already apply microorganism-based technology across the farm-to-food spectrum. For one thing, Novozymes inoculates seeds with microorganisms that take up nitrogen to form the root nodules. "So, we get much better growth with less dependence on chemical fertilizers," says Claus Crone Fuglsang, chief scientific officer at Novozymes. The company also makes biopesticides from biodegradable proteins derived from microorganisms.

Microorganism-derived enzymes from Novozymes can also be used to improve the taste, texture, and nutritional content of plant-derived foods. For instance, soy proteins lack some of the amino-acid combinations that make whey protein more nutritious. Instead of growing so many soybeans, a fermented protein can be engineered to contain the right amino acids for complete nutrition, while also providing a meaty flavor and texture. So, the production of plant-based foods such as meatless burgers can ease the pressure on the environment and improve human health. As Nielsen emphasizes, "Plant-based foods are more healthy and more sustainable in terms of their production."

Fuglsang calls Novozymes' combination of agricultural technologies "farm to fork." "It covers everything from growing crops to processing grains, from producing alternative proteins to improving taste, texture, and nutritional profiles of plant-based consumer products," he says.

Cleaning up the environment

Many other consumer products could benefit from microorganism-derived enzymes, which could also reduce their impact on the environment. One that Fuglsang points out is laundry detergent. "People don't think much about it, but laundry detergents can contain pretty harsh chemicals that get washed out to the environment," he says. "The right enzyme can replace some of these chemicals, help clean the clothes, degrade into harmless byproducts, and even allow a lower wash temperature."

PHOTO: © BLUEDOG STUDIO/SHUTTERSTOCK.COM

For Fuglsang, though, even more exciting advances lie ahead, and work on them is already underway. One that he mentions is recycling polyethylene terephthalate (PET), which is incorporated into many plastic containers, such as water and soda bottles. Novozymes and Carbios in France are running a pilot project in which enzymes break down PET into its components, which can then be used to make new plastic containers.

Another Earth-friendly project underway at Novozymes uses enzymes to capture CO_2 before it raises the levels of greenhouse gases. Novozymes is working with Italy-based Saipem to deliver enzyme-based solutions that can "reduce the amount of chemicals and heat that are needed, compared to other carbon capture systems," Fuglsang says.



Making better beer with microorganisms

Biotechnology can even help the beverage industry improve beer production. In Copenhagen, for example, scientists at EvodiaBio produce "natural and sustainable aromas that can bring more flavor and excitement to nonalcoholic beer," says CEO Camilla Fenneberg.

The technology behind EvodiaBio's methods emerged from Simon Dusséaux's postdoctoral work at the University of Copenhagen. In searching for more sustainable ways to produce molecules traditionally grown in plants, such as the hops for beer, Dusséaux—now cofounder and chief scientific officer at EvodiaBio—turned to fermentation. "The molecules are present in such low amounts in the plant material that you need to grow a lot of plants to just extract a few milligrams of what is of interest," he says. That approach is resource-intensive, especially for fuel and water.

After years of research, Dusséaux and his academic colleagues manipulated the genes in yeast to make large quantities of monoterpenes, which are flavor and aroma molecules "that can mimic the different profiles of various hops," Fenneberg says. EvodiaBio developed commercial processes to produce and extract the monoterpenes.

Although nonalcoholic beer offers many advantages, good taste is not usually one of them. So, EvodiaBio is starting in that market. As Fenneberg says, adding its molecules to a nonalcoholic beer creates a beverage with a "very nice hoppy flavor, like a regular beer." Even a professional beer-tasting panel agreed.

Although their beer is not yet on the market, the company has thus far deftly avoided the "Valley of Death"—the time from starting up until making money—through careful planning and help from BII. "I think what is really important is to ensure that you have a proven business model or reputation of your product before you invest in, for example, equipment and so on," Fenneberg says. "Plus, we went out very early to breweries and tested our product with them." She adds that it's important to surround yourself with experts in all of the relevant fields, which BII provided.

> **"I think everyone knows that the planet's current health is not very strong."**
> **—Jens Nielsen**

With ongoing innovation, biotechnology might solve some of Earth's biggest sustainability challenges, such as shrinking the Great Pacific Garbage Patch. "Through microbial fermentation, you can make polylactic acid, which is biodegradable and can be used in plastic bottles and bags," Nielsen says. Instead of ending up floating in the ocean, the plastic would be broken down and taken up by microorganisms. As Nielsen says, "That will be fantastic."

Sponsored by

BII BioInnovation Institute

HONORING INNOVATION

In 2021, BII and *Science* teamed up to create the BioInnovation Institute & *Science* Prize for Innovation. This prize honors researchers who turn life-science studies into products and commercial innovations. Consequently, the prize recognizes researchers who innovate within academics, but with an eye toward benefiting society. The submissions for the 2023 prize will open on 15 June 2022. Go to: science.org/bii

PHOTO: PROVIDED BY BII

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 Stella M. Hurtley (UK), Phillip D. Szuromi, Sacha Vignieri **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), L. Bryan Ray, H. Jesse Smith, Keith T. Smith (UK), Jelena Stajic, Peter Stern (UK), Valerie B. Thompson, Brad Wible, Yuen Yiu **ASSOCIATE EDITORS** Michael A. Funk, Bianca Lopez, Seth Thomas Scanlon (UK), Yury V. Suleymanov, Ekeoma Uzogara **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, 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** Tess Joose **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 **GRAPHICS MANAGING EDITOR** Chris Bickel **INTERIM MULTIMEDIA MANAGING PRODUCER** Sarah Crespi **PHOTOGRAPHY MANAGING EDITOR** William Douthitt **WEB STRATEGY MANAGER** Kara Estelle-Powers **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 **VIDEO PRODUCER** Meagan Cantwell **SENIOR PHOTO EDITOR** Emily Petersen **PHOTO EDITOR** Kaitlyn Dolan **SOCIAL MEDIA STRATEGIST** Jessica Hubbard **SOCIAL MEDIA PRODUCER** Sabrina Jenkins **WEB DESIGNER** Jennie Pajerowski

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 ANALYSTS** Kurt Ennis, Maggie Clark **FINANCIAL ANALYST** Isacco Fusi **BUSINESS OPERATIONS ADMINISTRATOR** Taylor Fisher **DIGITAL/PRINT STRATEGY MANAGER** Jason Hillman **SENIOR MANAGER, PUBLISHING AND CONTENT SYSTEMS** Marcus Spiegler **ASSISTANT MANAGER DIGITAL/PRINT** Rebecca Doshi **SENIOR CONTENT & PUBLISHING SYSTEMS SPECIALIST** Jacob Hedrick **SENIOR CONTENT SPECIALIST** 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** Aaron Helmbrecht, Ashley Hylton, Mike Romano, Tori Velasquez, Jenna Voris **SENIOR DESIGNER** Kim Huynh

DIRECTOR AND SENIOR EDITOR, CUSTOM PUBLISHING Sean Sanders **ASSISTANT EDITOR, CUSTOM PUBLISHING** Jackie Oberst **PROJECT MANAGER** Melissa Collins

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** Caroline Breul, 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 **SYSTEMS & OPERATIONS ANALYST** Ben Teincuff **FULFILLMENT ANALYST** Aminta Reyes

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 PERMISSIONS ASSOCIATE** Elizabeth Sandler **LICENSING ASSOCIATE** Virginia Warren **CONTRACT SUPPORT SPECIALIST** Michael Wheeler

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.*
James Analytis, *UC Berkeley*
Trevor Archer, *NIEHS, NIH*
Paola Arlotta, *Harvard U.*
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*
Malcolm Brenner, *Baylor Coll. of Med.*
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*
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*
Christina Curtis, *Stanford U.*
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*
Samuel Díaz-Muñoz, *UC Davis*
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*
Nader Engheta, *UPenn*
Karen Ersche, *U. of Cambridge*
Beate Escher, *UFZ & U. of Tübingen*
Barry Everitt, *U. of Cambridge*
Vanessa Ezenwa, *U. of Georgia*
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.*
Daniel Haber, *Mass. General Hos.*
Sharon Hammes-Schiffer, *Yale U.*
Wolf-Dietrich Hardt, *ETH Zürich*
Louise Harra, *U. Coll. London*
Carl-Philipp Heisenberg, *IST Austria*
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*
Gwyneth Ingram, *ENS Lyon*
Darrell Irvine, *MIT*
Akiko Iwasaki, *Yale U.*
Stephen Jackson, *USGS & U. of Arizona*
Erich Jarvis, *Rockefeller U.*
Peter Jonas, *IST Austria*
Johanna Joyce, *U. de Lausanne*
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*
Laura Lackner, *Northwestern U.*
Gabriel Lander, *Scripps Res.* **(S)**
Mitchell A. Lazar, *UPenn*
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*
Ari Pekka Mähönen, *U. of Helsinki*
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*
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*
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*
Judith Varner, *UC San Diego*
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.*
Li Wu, *Tsinghua U.*
Wei Xie, *Tsinghua U.*
Benjamin Youngblood, *St. Jude*
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*

2023 Vetlesen Prize

Achievement in the Earth Sciences

The Vetlesen Prize was established in 1959 by the G. Unger Vetlesen Foundation to honor scientific achievement that has resulted in a clearer understanding of the Earth, its history, or its relation to the universe. The prize, which is administered by Columbia University's Lamont-Doherty Earth Observatory, consists of a gold medal and a cash award of $250,000 and will next be awarded in 2023.

Nomination packages should include at least two letters that describe the nominee's contributions to a fuller understanding of the workings of our planet, a one-paragraph biographical sketch, and a full curriculum vitae of the candidate.

Nominations should be sent prior to **June 30, 2022** to: **vetlesenprize@ldeo.columbia.edu**

or via mail to: **Dr. Maureen E. Raymo**, Director, Lamont-Doherty Earth Observatory
PO Box 1000, 61 Route 9W, Palisades, NY 10964

For more information about the Vetlesen Prize visit **www.ldeo.columbia.edu/the-vetlesen-prize**

Columbia Climate School
Lamont-Doherty Earth Observatory

Past Vetlesen Laureates

2020	Anny Cazenave
2017	Mark A. Cane, S. George Philander
2015	Sir R. Stephen J. Sparks
2012	Susan Solomon, Jean Jouzel
2008	Walter Alvarez
2004	W. Richard Peltier, Sir Nicholas J. Shackleton
2000	W. Jason Morgan, Walter C. Pitman III, Lynn R. Sykes
1996	Robert E. Dickinson, John Imbrie
1993	Walter H. Munk
1987	Wallace S. Broecker, Harmon Craig
1981	M. King Hubbert
1978	J. Tuzo Wilson
1974	Chaim L. Pekeris
1973	William A. Fowler
1970	Allan V. Cox, Richard R. Doell, S. Keith Runcorn
1968	Francis Birch, Sir Edward Bullard
1966	Jan Hendrik Oort
1964	Pentti E. Eskola, Arthur Holmes
1962	Sir Harold Jeffreys, Felix Andries Vening Meinesz
1960	W. Maurice Ewing

NEWS FROM Science | AAAS

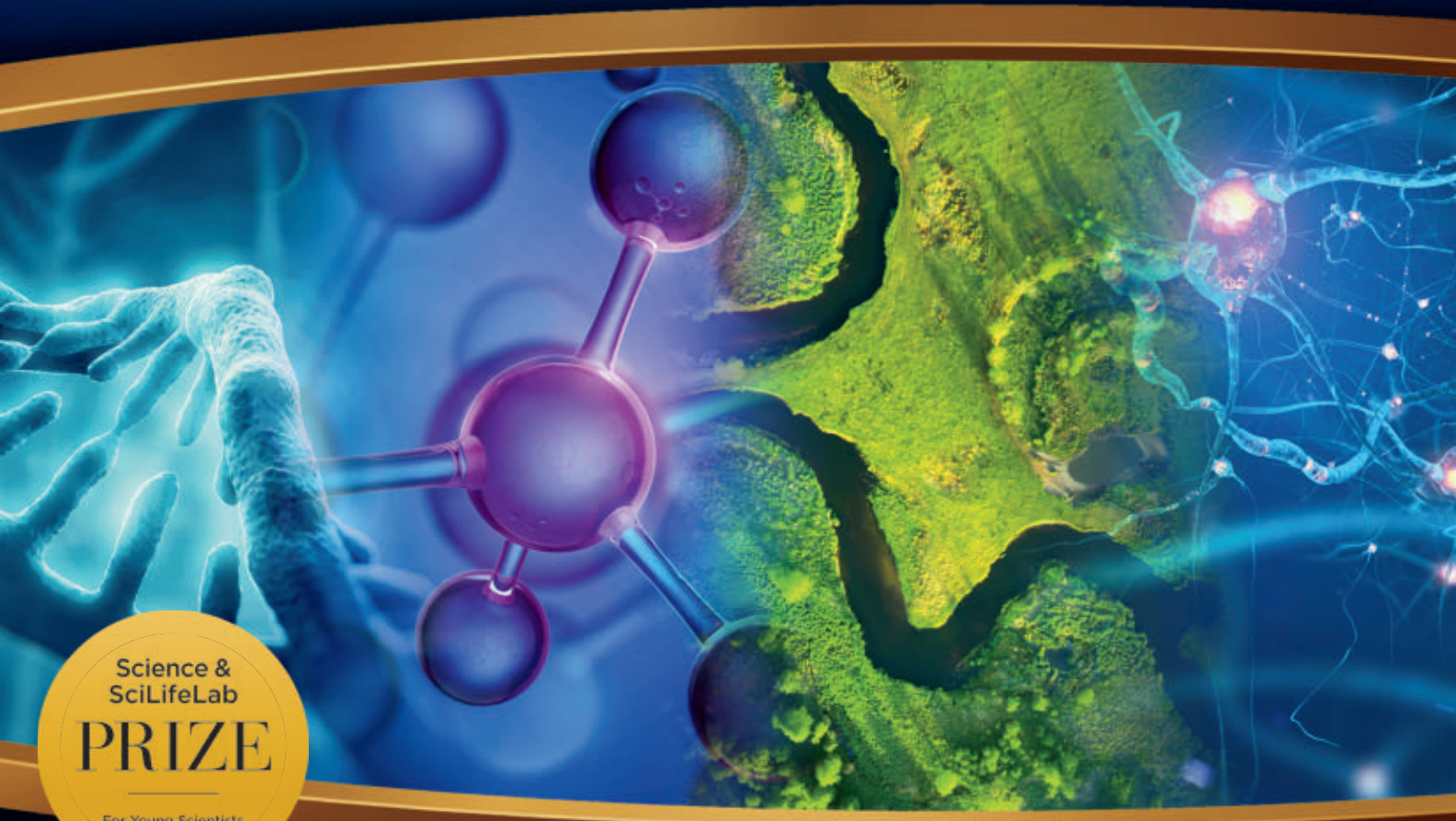
Subscribe to **News from Science** for unlimited access to authoritative, up-to-the-minute news on research and science policy.



bit.ly/NewsFromScience

Congratulations on your Ph.D.!
Ready for your next step?

Would you like to see your research published in *Science*, have the opportunity to receive USD 30,000, and get to travel to Sweden in December?



The *Science* & SciLifeLab Prize for Young Scientists is an annual prize awarded to early-career scientists. The prize is awarded in four research categories:

- Cell and Molecular Biology
- Ecology and Environment
- Molecular Medicine
- Genomics, Proteomics, and Systems Biology

Entrants for the 2022 prize must have received their Ph.D. between January 1, 2020 and December 31, 2021.

Applicants will submit a 1,000-word essay that is judged by an independent editorial team organized by the journal *Science*. Essays are assessed on the quality of research and the applicants' ability to articulate how their work would contribute to their scientific field.
Apply before July 15, 2022.

Knut och Alice Wallenbergs Stiftelse

SciLifeLab

Science AAAS

scienceprize.scilifelab.se

EDITORIAL

Faculty must lead inclusion

Later this year, I will step down as president of the University of Maryland, Baltimore County (UMBC), after 30 years. I know this to be true—faculty members have the power. Presidents, provosts, and senior leaders at an institution set a tone and create the setting for successful programs. But ultimately, it is the commitment of professors and teaching staff that determines whether students—all students—can pursue their interests and achieve their goals. It takes high expectations for both students and educators.

My own passion for helping people from all backgrounds to become scientists began in my childhood. I joined the 1963 Children's March in Birmingham, Alabama, because I wanted a better education. Along with hundreds of other Black youth who joined the protest, I spent nearly a week in jail. A year later, I participated in a federally funded program for high-achieving Black high school students. A math professor challenged us to work together on a difficult problem. Solving it took us 2 days. I learned from both experiences that high expectations and group work are powerful.

A quarter-century later, in the fall of 1989, UMBC decided to commit to diversity and inclusion in science and engineering by launching its Meyerhoff Scholars Program, which focused on Black students in science, technology, engineering, and mathematics (STEM) fields. (Later, the program would include students of all backgrounds.) Faculty pulled students into their research, and the university created settings in which students could work together, learning, growing, acquiring agency, and finding a sense of belonging. UMBC has used a holistic approach to admissions for its Meyerhoff program, considering test scores, coursework, and grades, but also a student's grit, drive, "fire in the belly," and excitement about STEM research.

Thirty years later, Meyerhoff Scholars are five times more likely to graduate from a STEM PhD or MD-PhD program than students who were invited to join the program but attended another university. UMBC is the number one producer of African American undergraduates who go on to earn PhDs in the natural sciences and engineering or an MD-PhD. Today, The Pennsylvania State University and the University of North Carolina, Chapel Hill, each have a Meyerhoff-like program. The Howard Hughes Medical Institute intends to fund 24 research universities to develop and implement similar programs. The Chan Zuckerberg Initiative is also funding programs at the University of California's Berkeley and San Diego campuses.

For university leaders to achieve long-term sustainable change in diversity and inclusion, they need to bring faculty into this work as allies in a "high expectation" strategy. Staff members offer crucial support, providing advice, professional development, and programming. Faculty create an empowering culture in which students find a sense of community and help each other study and learn. Because of this, university leaders want faculty to rethink the way they teach, redesigning courses to emphasize active and team-based learning, for example. University leaders want them to advise and mentor students and bring them into their research—in the lab or in the field. It takes scientists to produce scientists.

> **"When faculty take ownership of inclusion... universities will finally 'move the needle'..."**

Introducing this kind of "Meyerhoff change" in a university begins by collecting and analyzing data for science and engineering students. What percentage, by race and gender, is passing introductory biology and chemistry courses? Are they progressing to the next course and doing well? Are students persisting in STEM majors? Working with faculty, leaders can discuss information to understand problems, brainstorm solutions, and agree on a plan. Educators who have had success with underrepresented minority students should be invited to speak with the faculty. They can provide insight and inspire faculty members to become allies and champions by taking a lead in changing STEM education. Minority students can be invited to speak about their experiences as well. To make all of this work, institutional funds will be necessary, as will a commitment to raising money from agencies, corporations, foundations, and individuals.

It takes broad institutional commitment to produce scientists. Leaders set the tone, staff provide support, and faculty lead in the classroom and in the lab. When faculty take ownership of inclusion in science and engineering, mentoring undergraduates and graduate students, bringing them into the work, and championing them on their career journeys, then universities will finally "move the needle" on diversity in these fields.

–Freeman A. Hrabowski III



PHOTO: COURTESY OF FREEMAN HRABOWSKI

Freeman A. Hrabowski III
is president of the University of Maryland, Baltimore County, Baltimore, MD, USA.
hrabowski@umbc.edu

10.1126/science.abq7596

Cottrell Scholars 2022: the Future *of* Science

Research Corporation for Science Advancement, America's first foundation dedicated wholly to science, has named 24 early career scholars in chemistry, physics, and astronomy as its 2022 Cottrell Scholars. Each awardee receives $100,000. Recipients are recognized for high-quality research, innovative educational initiatives, and academic leadership.

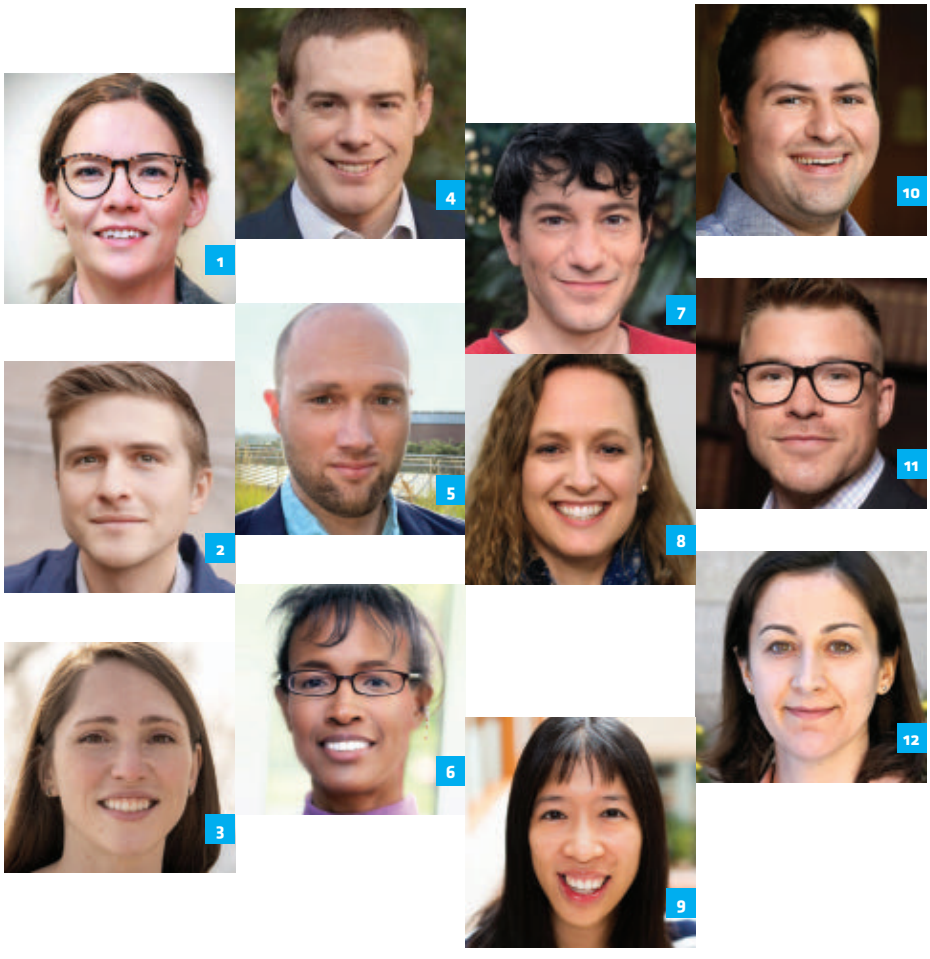
Three accomplished Cottrell Scholars (far right) further along in their careers have also received recognition for their achievements. STAR Awards reward excellence in science teaching and research. The IMPACT Award is given for national impact in science through leadership and service.

For additional information visit rescorp.org or call 520.571.1111.

Cottrell Scholar Awardees

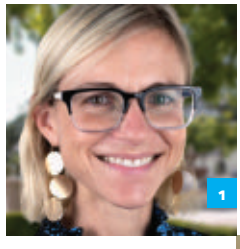
1. Darcy Barron, physics, University of New Mexico / Enabling Discoveries in Fundamental Physics by Maximizing the Sensitivity of Cosmic Microwave Background Polarization Surveys from Chile **2. Carl Brozek**, chemistry, University of Oregon / Clean Water from Porous Nanocrystals: An Undergraduate Training Program in Soft Skills and Sustainable Materials **3. Ilse Cleeves**, astronomy, University of Virginia / Identifying Molecular Patterns that Reveal the Chemistry of Planet Formation **4. Scott Cushing**, chemistry, California Institute of Technology / The Role of Picosecond Correlations in Solid-State Electrolytes for Batteries **5. Ryan Davis**, chemistry, Trinity University / Chemistry Beyond the Beaker: Exploring Supramolecular Assembly in Aqueous Microdroplets and Addressing Inequities in Chemical Education **6. Serena Eley**, physics, Colorado School of Mines / Identifying the Microscopic Origins of Energy Loss Mechanisms in Superconducting Quantum Circuits through Defect Landscape Engineering

7. Ben Feldman, physics, Stanford University / Thermodynamic Characterization of Metal-insulator Transitions in Dual Gated Moiré Superlattices **8. Kate Follette**, astronomy, Amherst College / Moving Forward—Towards Accurate Recovery and Interpretation of Accreting Protoplanets and a Socially Just Undergraduate Astronomy Curriculum **9. Wen-fai Fong**, astronomy, Northwestern University / Toward the Next Breakthroughs in Time-Domain Astronomy: The Origins of Fast Radio Bursts **10. Stephen Fried**, chemistry, Johns Hopkins University / Bringing New Life to Prebiotic Peptide Chemistry and to the Physical Chemistry Curriculum **11. Ryan Hadt**, chemistry, California Institute of Technology / Research—Learning How to Engineer Spin-Phonon Coupling in Molecules and Materials; Educational—Tackling Theoretical Topics in Inorganic Chemistry: A Worked Example Approach **12. Christine Hagan**, chemistry, College of the Holy Cross / Mechanistic Studies of Protein Toxin Delivery by Bacterial Contact-Dependent Inhibition Systems

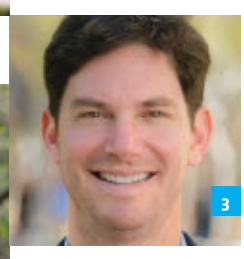


STAR and IMPACT Awardees

1. STAR awardee Rae Robertson-Anderson, physics, University of San Diego / Cottrell Scholar 2010
2. STAR awardee Scott Snyder, chemistry, University of Chicago / Cottrell Scholar 2009
3. IMPACT awardee Seth Cohen, chemistry, University of California, San Diego / Cottrell Scholar 2004

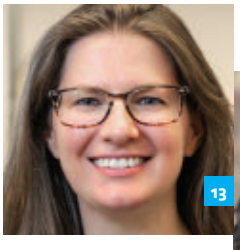


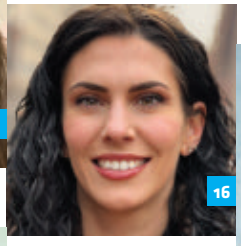


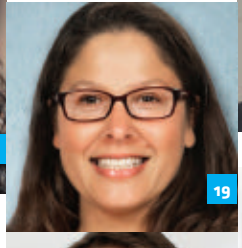


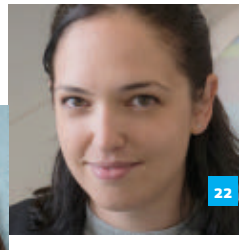
13. Sarah Keane, chemistry, University of Michigan / RNA Matchmaker: The Role of Loops and Mismatches in RNA Processing
14. Daniel Keedy, chemistry, CUNY Advanced Science Research Center & City College of New York / Illuminating Structural Motions that Underlie Allostery in Dynamic Phosphatase Enzymes **15. Michael Larsen**, chemistry, Western Washington University / Diverse N-Functionalized Polyureas by Cationic Ring-Opening Polymerization of Iminooxazolidines **16. Lauren Marbella**, chemistry, Columbia University / Tracking (Electro)Chemical Reduction at Electrode/Electrolyte Interfaces with Operando NMR
17. Krystle McLaughlin, chemistry, Vassar College / Structural Basis for the Conjugative Spread of Antibiotic Resistance
18. Jorge Muñoz, physics, University of Texas at El Paso / Fast and Accurate Computation of Anharmonic Phonons in Polymorphic Materials

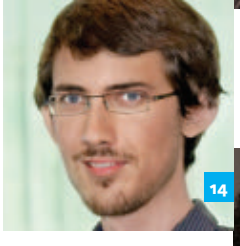
19. Lisa Olshansky, chemistry, University of Illinois at Urbana-Champaign / Exploring and Exploiting Conformational Dynamics for Proton-Coupled Electron Transfer
20. Zachariah Page, chemistry, University of Texas at Austin / Color-Coded Orthogonal Photochemistry from a Single Dye & Guiding Student Mindsets in Organic Chemistry
21. Amanda Patrick, chemistry, Mississippi State University / Winnowing the Possible Identities of Metabolomics "Features" by Hydrogen-Deuterium Exchange Mass Spectrometry **22. Orit Peleg**, physics, University of Colorado Boulder / The Physics of Firefly Communications: Principles and Predictions **23. Aurora Pribram-Jones**, chemistry, University of California, Merced / Reframing Interaction in Quantum Mechanical Ensembles and Across Chemistry Learning Communities **24. Ryan Trainor**, astronomy, Franklin & Marshall College / Feedback in the Circumgalactic Medium Probed with Lyman-alpha Emission

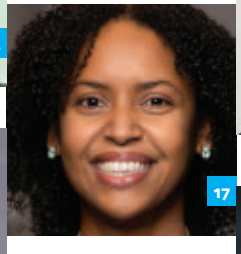


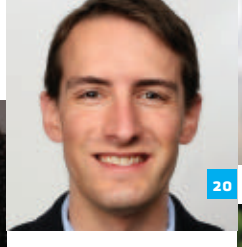


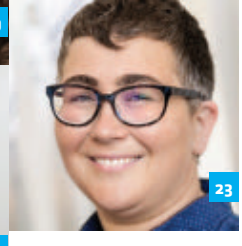


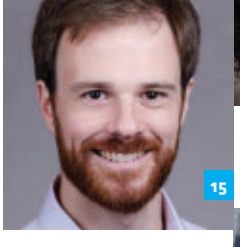


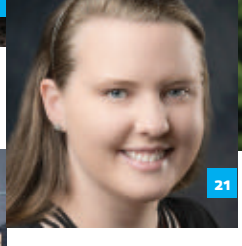


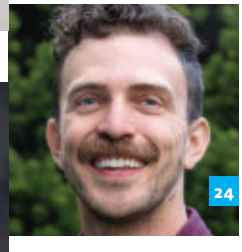


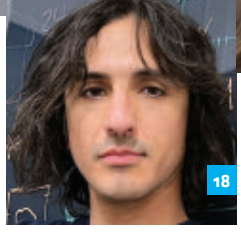












RESEARCH CORPORATION
for SCIENCE ADVANCEMENT
A foundation dedicated to science since 1912.

NEWS

> “It would be a huge shame for science and the public if this disappeared into the basement of an oligarch.”
>
> **Paleontologist Stephen Brusatte,** in *The New York Times,* about the first-ever auction on 12 May of a *Deinonychus* velociraptor fossil. He fears a private buyer might refuse to allow study or public display.



IN BRIEF

Edited by
Jeffrey Brainard

#METOO

Biologist accused of sexual misconduct drops NYU job quest

Cancer biologist David Sabatini, dogged by controversy since *Science* reported that he was in job talks with New York University’s Grossman School of Medicine, withdrew this week from consideration for a position at the medical school. Sabatini had recently been forced out of jobs at the Massachusetts Institute of Technology for breaching its consensual relationships policy and the Whitehead Institute for Biomedical Research for sexual harassment. The Howard Hughes Medical Institute also fired him last year for violating workplace policies. In a 3 May statement, Sabatini decried “false, distorted, and preposterous allegations about me” in news reports and on social media, adding that his withdrawal was meant to reduce “enormous pressure” on the medical school. Its top administrators emailed faculty members and students on 3 May that the school and Sabatini “have reached the conclusion that it will not be possible for him to become a member of our faculty.” Last week, hundreds of trainees protested outside the medical school (above), and hundreds of its faculty, students, alumni, and others signed letters of protest. Forty-five anonymous former Sabatini lab members said they signed a letter supporting him.

Probe knocks Raoult institute

RESEARCH MISCONDUCT | France’s drug regulatory agency has found “serious ethical breaches” in past clinical trials at the institute led by controversial microbiologist Didier Raoult, who became notorious for promoting hydroxychloroquine as a COVID-19 drug despite a lack of evidence. The National Agency for the Safety of Medicines and Health Products (ANSM) investigated two trials at the Hospital Institute of Marseille Mediterranean Infection (IHU), which Raoult leads, one on children’s infections with the bacterium *Tropheryma whipplei*, the other on illnesses contracted by French medical students while abroad. ANSM’s draft report, published on 27 April, documents “critical” and “major” problems in consent and ethical approval procedures in both studies and says Raoult submitted a falsified ethical approval document in the course of the investigation. ANSM has referred its findings to Marseille authorities for possible criminal prosecution. The report is another step in Raoult’s fall from grace, following a reprimand last year from

PHOTO: ABBY WILSON/WSN

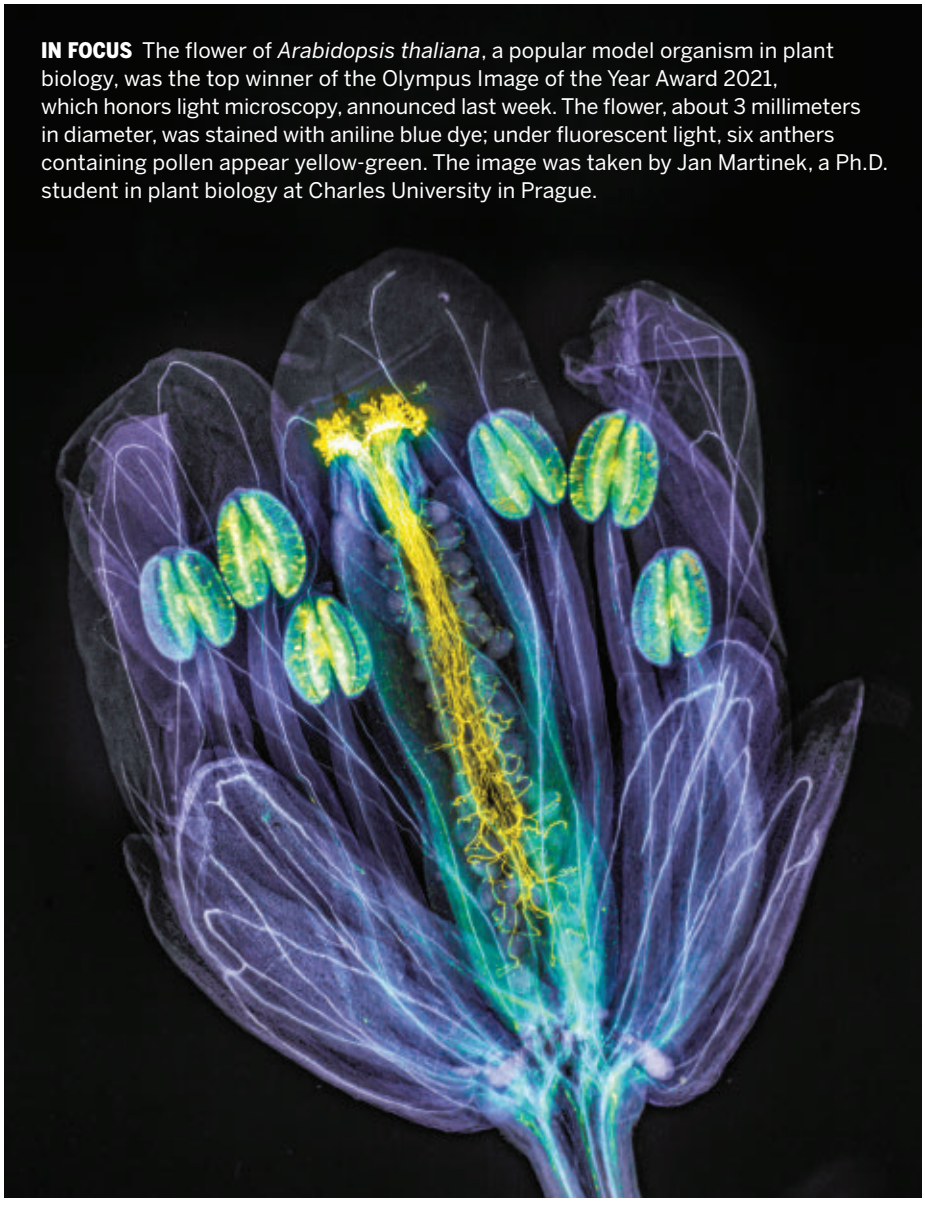
the French Medical Council for his hydroxychloroquine advocacy. ANSM's investigation will continue to scrutinize further trials, alongside investigations by other French authorities. In a statement to *Science*, Raoult said IHU followed ethical procedures and that the investigation may be an attempt to undermine his credibility, given his role as a witness in an investigation into the government's COVID-19 management.

COVID-19 pills fail at prevention

CLINICAL TRIALS | Pfizer's antiviral pill combination fails to prevent infection among household contacts of patients with COVID-19, a clinical trial has found. Paxlovid, a mix of two drugs, nirmatrelvir and ritonavir, is prescribed to COVID-19 patients with risk factors such as older age and obesity that place them at higher risk of hospitalization or death. Now, Pfizer has tested it in nearly 3000 household contacts of people who tested positive for the coronavirus. The trial found that when adults who had an infected household member took the drug for either 5 or 10 days, they were 32% or 37% less likely than those on a placebo to later test positive themselves—but the differences between the groups were not statistically significant, the company said, and could have resulted from chance. Some antivirals, such as those for HIV, have worked as preventives.

A bid to recruit Russian scientists

WAR | A U.S. proposal to poach Russian researchers opposed to the invasion of Ukraine is receiving mixed reviews. Last week, Bloomberg and other media outlets reported that President Joe Biden's administration, as part of a request to Congress for $33 billion in spending on Ukraine, is seeking to ease visa requirements for the next 4 years for Russian scientists in high-tech fields such as artificial intelligence, quantum physics, and semiconductors. Some administration officials compare the concept to Operation Paperclip, a clandestine U.S. effort after World War II that spirited more than 1600 researchers—many of them rocket scientists—out of Germany before the Soviet Union could absorb them. The overall spending measure has drawn bipartisan support in Congress. But, "It's really crazy" that the new proposal is being discussed openly beforehand, says Gerson Sher, an independent consultant and expert on Russian science engagement. Russia, he notes, could now move to cloister top scientists or otherwise restrict them from traveling abroad.



IN FOCUS The flower of *Arabidopsis thaliana*, a popular model organism in plant biology, was the top winner of the Olympus Image of the Year Award 2021, which honors light microscopy, announced last week. The flower, about 3 millimeters in diameter, was stained with aniline blue dye; under fluorescent light, six anthers containing pollen appear yellow-green. The image was taken by Jan Martinek, a Ph.D. student in plant biology at Charles University in Prague.

PHOTO: JAN MARTINEK/OLYMPUS IMAGE OF THE YEAR AWARD 2021

China tech giant boosts research

FUNDING | Tencent, the Chinese social media, e-commerce, and video game conglomerate, will put up $1.5 billion over 10 years to support basic research in biology, math, medicine, and physics in that country, Chinese media reported last week. The New Cornerstone Researcher Project is expected to allow up to 300 researchers to freely pursue their scientific interests. The new fund is one of several Tencent philanthropic efforts responding to government calls for successful high-tech companies to share their prosperity with society. Private and corporate philanthropy has been growing in China but the focus has long been on education and poverty alleviation, although the COVID-19 crisis prompted donations for research on vaccines and antiviral drugs. Tencent's move aligns with a government drive to boost support for basic research from 6% of the country's total R&D spending in 2021 to 8% by 2025. The fund will begin accepting applications for grants later this year.

UC campus buys Venter building

BIOLOGY | The nonprofit research center founded by renowned geneticist Craig Venter has sold its laboratory building in San Diego, California, to the University of California, San Diego (UCSD), for $25 million. The transaction occurred in January, and the J. Craig Venter Institute (JCVI) announced it last week. The sale allows JCVI to operate rent-free there for

ASTRONOMY

NASA grounds airborne telescope over cost concerns

NASA and its partner the German Space Agency said last week they will shut down the Stratospheric Observatory for Infrared Astronomy (SOFIA, pictured), a 2.5-meter telescope mounted in a Boeing 747 that flies above infrared-absorbing layers of the atmosphere. Working since 2014 to observe planets, star-forming regions, and nearby galaxies, SOFIA discovered water on the sunlit surface of the Moon in 2020. But astronomers said its high operating cost—$85 million annually, close to that of the Hubble Space Telescope—was not justified by its scientific output. The 2021 decadal survey in astrophysics, a high-level community report, recommended its termination. Defenders say it is the only instrument that can observe at far-infrared wavelengths. Congress has thwarted NASA's previous attempts to ax SOFIA.



5 years; the institute also retains a separate location in Rockville, Maryland. Many of JCVI's scientists already held joint appointments at UCSD, and JCVI is considering a follow-up deal to transfer some of its research programs to the university, the institute said in a press release. The sale came after JCVI's revenue had dropped by nearly half from 2018 to 2019, to $23.8 million, according to its federal tax filing for 2019. Selling the building "was an opportunity to free up resources to continue the important research programs that Craig and JCVI are doing and want to do," a spokesperson said. JCVI was created in 2006 from the merger of the Institute for Genomic Research, which Venter founded in 1992, and other Venter-backed research organizations. Among JCVI's genetics accomplishments is creating the first cell with a completely synthetic genome. Before that, Venter played a key role in sequencing the human genome.

Fund minority schools, panel says

DIVERSITY | Black colleges need more money, not just kind words, from the U.S. Department of Defense (DOD), an expert panel says. Historically Black colleges and universities (HBCUs) received just 0.4% of DOD's $6.6 billion annual budget for basic research in 2019, according to a 28 April report from the National Academies of Sciences, Engineering, and Medicine. "There is a clear disconnect between the expressed encouragement by Congress and DOD to increase participation of [minority-serving institutions] in defense-funded research and the resources allocated," the report says. It notes that NASA and the Department of Energy give HBCUs a much bigger slice—0.9%—of their research budgets. This year, Congress appropriated $100 million, double DOD's request, for a program that targets institutions enrolling large numbers of underrepresented students. The report says DOD officials should be asking for at least that amount each year for the program.

Diabetes compound cuts weight

BIOMEDICINE | An experimental diabetes drug used to also treat obesity helped people lose up to 22.5% of their body weight over 16 months, according to unpublished data released last week by the drug's manufacturer. In a press release, Eli Lilly & Company said about 630 people who gave themselves a weekly, 15-milligram injection of the drug, tirzepatide, lost 23 kilograms on average. Others taking lower doses lost less weight, and those receiving a placebo lost only 2 kilograms. Tirzepatide, developed for people with type 2 diabetes, mimics two gut hormones—glucagon-like peptide-1 (GLP-1) and gastric inhibitory polypeptide—that curb appetite and slow stomach emptying, which helps people eat less. Another GLP-1–mimicking drug, Novo Nordisk's semaglutide (Wegovy), approved by the U.S. Food and Drug Administration in June 2021, helped people lose 15% of their body weight over a similar period. Some specialists have called the two drugs important alternatives to bariatric surgery for obesity. One unknown is whether health insurers will cover the drugs; Wegovy's list price is $1350 a month.

New font of exotic nuclei opens

NUCLEAR PHYSICS | The U.S. Department of Energy this week officially opened the Facility for Rare Isotope Beams (FRIB) at Michigan State University. Researchers hope to use it to improve understanding of how heavier elements are forged in stellar explosions, among other topics. The $730 million linear accelerator can generate beams of any atomic nucleus from hydrogen to uranium. Physicists fire the beam through a graphite target to break the nuclei into lighter, exotic ones rich in neutrons. FRIB may produce 1000 kinds of nuclei never before observed.

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

PHOTO: NASA/JIM ROSS

eppendorf & Science

PRIZE FOR NEURO BIOLOGY



2021 Winner
Amber L. Alhadeff, Ph.D.
Monell Chemical Senses Center, USA
For research on the gut-brain control of hunger circuits

Tell the World About Your Work!

Application Deadline
June 15, 2022

Eppendorf & Science Prize for Neurobiology
The annual Eppendorf & Science Prize for Neurobiology is an international prize which honors young scientists for their outstanding contributions to neurobiological research based on methods of molecular and cell biology. The winner and finalists are selected by a committee of independent scientists, chaired by *Science's* Senior Editor, Dr. Peter Stern. If you are 35 years of age or younger and doing great research, now is the time to apply for this prize.

As the Grand Prize Winner, you could be next to receive
> Prize money of US$25,000
> Publication of your work in *Science*
> Full support to attend the Prize Ceremony held in conjunction with the Annual Meeting of the Society for Neuroscience in the USA
> 10-year AAAS membership and online subscription to *Science*
> Complimentary products worth US$1,000 from Eppendorf
> An invitation to visit Eppendorf in Hamburg, Germany

It's easy to apply! Write a 1,000-word essay and tell the world about your work. Learn more at:

eppendorf.com/prize

eppendorf Science AAAS

AAAS® and Science® are registered trademarks of the American Association for the Advancement of Science, USA. Eppendorf® and the Eppendorf Brand Design are registered trademarks of Eppendorf SE, Germany. All rights reserved, including graphics and images. Copyright © 2022 by Eppendorf SE. Photography: Happy Hour Headshot Philadelphia.



IN DEPTH

Peat layers in Poland's Śnieżka bog capture spikes in plutonium from nuclear tests, jumps in metals from nearby smokestacks, and the extinction of native amoebas.

GEOLOGY

Bids for Anthropocene's 'golden spike' emerge

Sites compete to mark global changes of the 1950s and define new geological age

By **Paul Voosen**

If you had to pick one spot that best reflects when human activity became an Earth-shaping force, where would it be? Geoscientists will consider the question this month when they meet to evaluate 12 sites, only one of which can serve as the "golden spike" for the Anthropocene, a proposed geological age beginning in the 1950s amid the fire of nuclear bomb tests and the fumes of surging fossil fuel use.

Although the idea of the Anthropocene has gained wide traction, it still lacks a formal geological definition. In 2016, the Anthropocene Working Group (AWG), a group of several dozen geoscientists convened by the International Commission on Stratigraphy (ICS), settled on the early 1950s as its starting point (*Science*, 26 August 2016, p. 852). But the ICS still needs a formal proposal with an ideal geologic sample recording these global changes—a golden spike—to mark the end of the Holocene epoch, which began 11,700 years ago, and the beginning of the Anthropocene.

To find that sample, teams of earth scientists spent several years analyzing sites that contain promising markers, such as spikes in plutonium and other radionuclides that settled after atmospheric nuclear tests, spherical ash particles from unchecked industrial emissions, microplastics, and perturbations to carbon and nitrogen chemistry from greenhouse gas emissions and urban smog. The array of possibilities was bewildering, says Colin Waters, a geologist at the University of Leicester and AWG chair. "Starting from scratch, you've got the whole world to play with."

At a meeting in Berlin, starting on 18 May, the teams will present the case for their site to serve as the golden spike. Months of deliberation will begin this summer, before a final silent voting period. If 60% of the group's members agree on one site, a selection will be announced by December, Waters says.

Some of the strongest candidates come from lake bottoms that accumulate muds in thin annual layers, creating high-resolution records. Crawford Lake in Canada's Ontario province, only some 200 meters wide but 25 meters deep, is one. In cores of mud from the lake, changes seen since the 1950s stand out vividly, including the bomb spike and an "off the charts" increase in soot from local industry, says Francine McCarthy, a micropaleontologist at Brock University. "We have a really ideal site," she says.

A site outside Europe or North America, which industrialized well ahead of other regions, might be a better record of changes that swept the world in the 1950s, says Yongming Han, a geochemist at the Chinese Academy of Sciences's Institute of Earth Environment. His team will propose Sihailongwan Lake, in a protected forest in northeastern China. Mud cores from the lake show not only the bomb spike and rises in soot, but also increases in spherical ash particles and heavy metals. They also record a rapid increase in light carbon, the isotope favored by life, driven by burning fossil fuels that were once organic matter.

There is no greater aggregator of global signals than the ocean, making marine sediments strong contenders. One set of muds, collected in the Baltic Sea, shows a pronounced change from light gray to dark brown in the 1950s, the result of algal blooms fed by fertilizer in farming runoff. "There's an explosion" of algae, says Jérôme Kaiser, a marine geochemist at the Leibniz Institute for Baltic Sea Research.

PHOTO: MAREK DOBROWOLSKI

His team sees not only the bomb spike and surge of spherical ash, but also a rise in DDT, a once-ubiquitous pesticide. Meanwhile, in Japan, sediment samples from Beppu Bay show an increase in total nitrogen from fertilizer runoff around this time, says Michinobu Kuwae, a paleoceanographer at Ehime University. "Humankind must notice that we are already in the Anthropocene," he says.

Corals are another ocean record under consideration, but they seem to be less suited for a golden spike, says Kristine DeLong, a paleoclimatologist at Louisiana State University, Baton Rouge. The coral she studied from a pristine reef in the Gulf of Mexico did not even capture the bomb spike. "It's kind of nice, these corals weren't impacted," she says. What they do display are strange increases in barium—likely from barite, a mineral used in offshore oil drilling.

A late addition to the slate of candidates came from a Polish peat bog near a mountain summit in a region known for its mines and heavy industry. The bog is fed only by rain, so it forms a great archive of atmospheric pollution, says Barbara Fiałkiewicz-Kozieł, a geoecologist at Adam Mickiewicz University. Its sediments capture the typical 1950s markers, along with spikes in lead, aluminum, titanium from nearby smokestacks—and even the extinction of amoebas that once thrived in the peat. "I have 30 different proxies," she says. "And every proxy gives a very distinct answer."

After the AWG settles on a site, the proposal must be approved by 60% of the two dozen or so members of the Subcommission on Quaternary Stratigraphy. From there it will go to the executive committee of the ICS—and finally to the International Union of Geological Sciences (IUGS). These are not rubber stamps; proposals for stratigraphic divisions routinely fail, and when they do, they cannot be reconsidered for 10 years. "There's no guarantee that the Anthropocene will be agreed upon," says Jan Zalasiewicz, a Leicester stratigrapher and longtime AWG head who now leads the quaternary subcommission. "There are influential stratigraphers who are very uncomfortable with a formalized Anthropocene."

One of those stratigraphers is IUGS head Stanley Finney of California State University, Long Beach. From the start, he says, the AWG has operated backward, starting with the proposed Anthropocene and looking for markers, rather than starting with the geologic record itself. The group "has not been open-minded," he says. In his view, the Anthropocene should not be shoehorned into the record of geologic time because, so far, it is just one human lifetime long, based on centimeters of mud and signals that may not persist once those sediments turn into rocks. Even the plutonium will be largely undetectable after 100,000 years.

Despite his personal opposition, Finney says he won't stand in the way if the ICS is open and follows procedure. But he would prefer to categorize the Anthropocene as an "event," an informal term geoscientists use for everything from gradual planet-wide changes that take millions of years to large asteroid impacts. Waters agrees the Anthropocene can be an event, but argues that shouldn't preclude a more formal designation. He points out how the dinosaur-killing Chicxulub impact 66 million years ago is an event—but also one that marks the geological divide between the Cretaceous and the Paleogene periods.

AWG member Erle Ellis, a landscape ecologist at the University of Maryland, Baltimore County, believes the odds of approval for an Anthropocene epoch aren't great, but says the "event" idea could fare better. For years, the ICS has fretted about provoking negative publicity if it rejects the Anthropocene proposal. With an "event," Ellis says, the ICS can still acknowledge the significant impact that humanity is making on Earth. "The Anthropocene is never going to be rejected," he says. "It's simply a matter of how it's going to be defined." ■

Marking time

By December, a winner could emerge from 12 sites competing to be the Anthropocene's "golden spike." But its adoption by the bureaucracy that governs geologic time is not guaranteed.

NAME	LOCATION	TYPE
Beppu Bay	Kyushu Island, Japan	Marine sediments
Crawford Lake	Ontario, Canada	Lake muds
Ernesto Cave	Italy	Cave deposits
Flinders Reef	Coral Sea, Australia	Coral
Gotland Basin	Baltic Sea	Marine sediments
Palmer ice core	Antarctic Peninsula	Ice sheet
San Francisco Estuary	California, United States	Marine sediments
Searsville Reservoir	California, United States	Lake muds
Sihailongwan Lake	Jilin province, China	Lake muds
Śnieżka bog	Poland	Peat layers
Vienna Museum excavation	Austria	Urban soil
West Flower Garden Bank	Gulf of Mexico	Coral

DEMOGRAPHY

Census aims for better U.S. statistical portrait

Agency wants to retool its surveys and decennial census to improve efficiency and generate better data

By **Jeffrey Mervis**

Fresh from a tumultuous 2020 census, the U.S. Census Bureau has quietly begun what it hopes will be a major overhaul of how it conducts the decennial census and manages 130 surveys that paint a statistical picture of the country.

Its goals are to hold down costs, reduce the burden on survey respondents, and do better at finding people from groups often undercounted. Agency officials believe the changes will result in more useful and timely information about the nation's residents and the economy.

An independent team of former census directors and prominent social scientists based at the University of Virginia (UVA) is helping the agency meet those goals, and last week it released a progress report on its 4-year effort. "The Census Bureau is already moving in the right direction, and we're hoping to help them get there by exploring some of the scientific issues that need to be addressed," says Sallie Keller, a UVA statistician leading the team along with former Census Director Kenneth Prewitt.

"We understand that we need to take a different approach for 2030," says Robert Santos, a statistician who began a 5-year term as Census director in January. The agency faces ever-rising costs (the 2020 census cost $14.2 billion), declining response rates, diminished public trust in government, and heightened concerns about privacy. "We're looking to see how far we can go with the vision that has been expressed by the University of Virginia team," Santos says.

The research community, which could gain easier access to a richer trove of data sets, is rooting for the Census Bureau to succeed. But many are skeptical that it can



Existing U.S. surveys provide an incomplete picture of elderly care facilities.

adopt meaningful improvements in today's hyperpartisan political climate.

The agency's constitutional duty—to determine how to divvy up the 435 seats in the U.S. House of Representatives by state—makes it a perennial target for politicians. The 2020 cycle was especially fraught as then-President Donald Trump waged an assault on the agency's independence that included an unsuccessful campaign to add a citizenship question to the census.

One major challenge for the agency is to reduce its reliance on a master list of addresses, painstakingly compiled and continuously updated, as the basis for the decennial head count. But counting every household requires an army of enumerators to make repeated follow-up visits to dwellings believed to house residents who did not respond to the initial census questionnaire.

The Census Bureau hopes to dramatically shrink that mop-up effort in 2030 by identifying most of the U.S. population through so-called administrative records: personal information residents have already provided to other federal agencies, as well as data collected by local and state governments and by commercial vendors. The bureau could then conduct a much smaller follow-up effort to track down the estimated 15% to 20% of the population lacking a usable electronic footprint.

There's no guarantee, however, that data culled from administrative records would meet the bureau's high standards. And enumerating people who are missing from those records is no trivial matter, says Joseph Hotz, a population economist at Duke University.

Hotz is conducting a study to understand rural aging in North Carolina and has found that men living alone can be especially elusive. "They aren't homeowners, and they don't appear in commercial housing databases," he says. "They aren't easy to find because they don't want to be found."

Another major goal is to make better use of data from the welter of surveys the bureau now conducts apart from the decennial census. Those surveys generate separate master lists, or frames, that contain data such as the number and location of housing units, a roster of employers and jobs, and demographic information for the population as a whole.

Combining those frames and analyzing the merged data, Santos says, "might answer policy questions you couldn't even put on the table before." It could also lead to shorter, less burdensome surveys, he adds, by eliminating redundant questions.

The UVA team is conducting a series of pilot studies to demonstrate the added value of what it calls a "curated data enterprise." Joe Salvo, former chief demographer for New York City and now a senior fellow at UVA, says one example of an important economic trend that current surveys have struggled to capture is the gig economy, in which people hold down multiple jobs.

The bureau's Current Population Survey (CPS), a monthly survey of the U.S. labor force, yields a gross undercount of gig workers by assuming respondents have just one primary job, Salvo asserts. "CPS says there are 15 to 18 million people in the gig economy, and that the number hasn't changed since 2005," he says. "But recent studies suggest the number could be as high as 55 million." He thinks a better answer would come from combining CPS data with Internal Revenue Service data on reported income, along with records from nonprofit organizations and from those who are self-employed.

Another UVA project, on the state of nursing homes and residential care facilities, aims to create a more detailed picture of those who live in, work at, and operate such facilities. Keller says that picture could be pieced together from geospatial data and business and labor surveys and would be invaluable for responding to a natural disaster or pandemic.

Despite the potentially huge payoff, such efforts raise the specter of the federal government as an all-knowing Big Brother. Some of those fears could be alleviated, Keller says, by pooling the data only as needed, to answer a specific policy question.

Another safeguard is to cloak personal, individualized data. "We're very close to having the technology for creating a firewall that allows access to the data without actually being able to look at individual records," Santos explains.

The Census Bureau is already prohibited from releasing any data sets that, when combined with other anonymous information, could allow outsiders to identify individuals. To meet that requirement, the 2020 census applied a mathematical tool called differential privacy, which injects statistical "noise" into the data.

However, many demographers have complained that the amount of noise needed to avoid disclosure has undermined data quality. They worry the problem will get worse if the Census Bureau sets too high a privacy bar for any data it releases. One proposed solution would be to exempt some data, such as addresses, that are typically available on the internet.

Privacy concerns aren't the only reason for the decline in survey participation, social scientists say. The public is also skeptical that surveys are worth the time. "It's insufficient to tell people they are required by law to participate," Santos says. "We need to demonstrate tangibly how the data we're collecting is benefiting them on a day-to-day basis."

Nancy Potok, former chief statistician for the United States and a former deputy Census director, wholeheartedly endorses what Santos and the UVA team are trying to accomplish. But she wonders whether the agency will be able to make headway.

"Any big changes will make the Census Bureau a [political] lightning rod," she predicts. "And Census tries hard to avoid that fate. So in the end, I think that the status quo … is the most likely outcome."

Prewitt hopes otherwise and is counting on the research community to help the agency make its case for change. "Our challenge is to defend something that's inherently political with something that isn't political," he argues. "And the answer is science. That's the only thing that will carry the day." ■

PHOTO: CRAIG F. WALKER/*THE BOSTON GLOBE*/GETTY IMAGES

BIOMEDICINE

Doubt cast on inflammation's stop signals

Critics challenge data underpinning "resolution immunology," triggering university probes

By **Gunjan Sinha**

For 3 decades, biochemist Charles Serhan at Brigham and Women's Hospital has been the torch bearer for "resolution immunology"—a research field he pioneered after discovering the first molecules that seem to tamp down, or resolve, inflammation, the body's rapid cellular and chemical response to injuries, infections, and even allergens. Ever since, research on these stop signals, formally known as specialized pro-resolving mediators (SPMs), has exploded (*Science*, 2 January 2015, p. 18).

Scientists around the world have implicated multiple classes of these lipid-based molecules in all manner of diseases. A search on PubMed for "lipoxin," the first SPM identified, turns up more than 2200 papers, and at least three biotech companies are planning clinical trials with synthetic molecules meant to restore or boost the body's natural ability to end inflammation, which can wreak havoc if prolonged.

But what if a large chunk of the research on SPMs is based on flawed science?

Last month, in a paper in *Frontiers in Pharmacology*, an international group of 18 scientists including specialists in lipid analysis argued that the studies identifying these molecules in people typically detect them at levels so low they are unlikely to play a major role in resolving inflammation. The critics also found that most studies of human samples merely correlate SPM levels with a phase during the arclike time course of inflammation; data proving SPMs actually resolve it are weak to nonexistent, the authors argue.

The paper's authors offered an even more damning critique: Protocols developed by Serhan and his collaborators to detect SPMs in body fluids or tissue samples don't conform to accepted standards, they say. What is provided as proof of SPMs in many papers should actually be interpreted as mere "noise," they argue. "You cannot ignore the fact that there is a growing number of analytical chemists who cannot detect these things," says Nils Helge Schebb, chair of food chemistry at the University of Wuppertal and lead author.

Even some scientists who didn't join the broadside have struggled. Immunology researcher Derek Gilroy of University College London, who has collaborated with Serhan but is now moving his lab away from SPMs, tells *Science* he has had a "very negative experience working with these things and seen some data that makes me question whether these things are real."

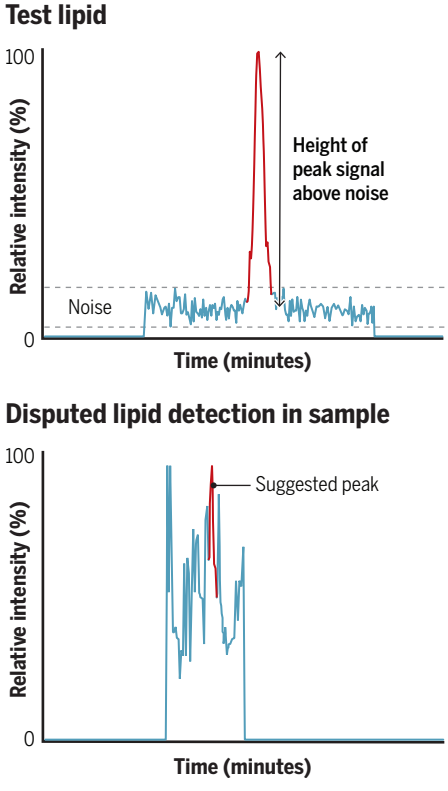
The concerns over the detection methods, *Science* has learned from several sources, have prompted investigations at Harvard Medical School, the parent organization of Serhan's hospital, and at Queen Mary University of London (QMUL), where molecular pharmacologist Jesmond Dalli, a former Serhan postdoc, now studies lipid mediators.

Neither university would confirm the probes to *Science*. The criticisms in the *Frontiers* commentary are "not true," Serhan said in an email. "Many others have indeed detected SPMs in human samples." A single critical paper compared with thousands on SPMs "is not a valid comparison of the facts. It's not clear to me what the goal of their 'review' is because it's totally wrong."

"I am baffled," adds QMUL immunopharmacologist Mauro Perretti, who studies SPMs independently of Dalli but also collaborates with Serhan. "SPMs exist, they have been characterized in multiple ways, they have been synthesized, and they are active." SPM proponents also note that animal studies and one human trial have already yielded hints that synthetic versions of the molecules work as advertised.

Point of contention

To be content a sample contains a molecule such as a lipid, scientists usually require a sharp peak, far above the noise, in data from liquid chromatography mass spectrometry (top). Critics of research on inflammation-resolving lipids say several chromatograms in one paper don't show a clear peak (bottom).



GRAPHIC: K. FRANKLIN/*SCIENCE*

SEARCHING FOR SIGNALS

Serhan has been studying inflammation since his postdoc years at the Karolinska Institute. In 1984, he discovered a novel fatty acid made by the body. Over time, he and his collaborators piled up evidence that the molecule, which they dubbed lipoxin, could shut down inflammation. At the time researchers believed inflammation just dissipated, like a crowd after an event. Serhan argued instead that lipoxin and other molecules actively end the response, like riot police arriving to disperse the crowd. Since then, his lab and others around the world have found and characterized more than 20 other types of lipid-based biochemicals that form four SPM classes—lipoxins, resolvins, maresins, and protectins.

Schebb began to look for SPMs about 6 years ago but had trouble detecting any of them in human samples. He visited Dalli's lab, looking for advice, but says "they didn't teach me anything that helped." Schebb soon learned that several other scientists also had trouble detecting SPMs, and he became increasingly skeptical of the research.

Enter Valerie O'Donnell, a lipid biochemist at Cardiff University. About 3 years ago, when reviewing papers on SPMs, O'Donnell noticed a few "strange" looking figures showing the results of liquid chromatography mass spectrometry (LCMS), which separates a sample's molecular contents into ions with distinct molecular masses, producing a chart called a chromatogram. These diagrams are supposed to show a sharp peak representing a detected molecule. But when O'Donnell looked at high-resolution images of

chromatograms said to show SPMs, the figures didn't look like "real data," she says.

She consulted Ian Blair and Garret FitzGerald, both at the University of Pennsylvania, and Robert Murphy at the University of Colorado, Denver, who also study lipids, and together they analyzed other published SPM papers. They found at least 70 papers between the Serhan and Dalli research groups with chromatograms that they considered suspicious.

They took some of their criticism public last year. In a preprint posted on 8 December 2021, the same group together with other specialists in lipid analysis highlighted a paper published by Dalli's group in 2020 that identified a lipoxin, a resolvin, and a maresin in the serum of people with early rheumatoid arthritis. The paper went on to suggest the SPMs could be used as biomarkers to assess patients' response to medications. But the critics said Dalli was not setting a limit of detection—a value representing the lowest measurable concentration that provides high confidence a molecule exists. Instead, they say, he used a different method that did not comply with accepted standards for detecting biomolecules. When O'Donnell and colleagues applied the criteria described in the paper to inert methanol and buffer fluid, they found it indicated the presence of lipids where clearly none existed.

O'Donnell and her colleagues concluded that Serhan and Dalli often appeared to quantify lipids by integrating what the critics would consider "noise." "We had never seen mass spectrometry quantitation performed this way before," Blair says.

In a rebuttal posted this week as a preprint on bioRxiv, Dalli, Serhan, and a colleague defend their methods by reanalyzing some of their LCMS data using the limit-of-detection criteria suggested in the preprint. The outcome confirms his group's original results, Dalli says. Dalli also notes that an analysis of blank samples on his lab's mass spectrometry instruments did not yield false signs of SPMs. Dalli canceled an interview after *Science* contacted QMUL about the dispute but answered questions by email, calling the conclusions of the preprint "incorrect." He added: "I really cannot understand why someone would feel they need to go to these lengths to attack me and my colleagues."

NO RESOLUTION

With the *Frontiers* paper, the critics have now broadened their case against the SPM work. But if they are right about the flaws in the analyses, how did these papers pass peer review and get published? FitzGerald says there are very few experts in mass spectrometry of low abundance lipids. Consequently, submitted papers about SPMs are usually sent just to experts in the particular disease being studied, who do not necessarily understand the intricacies of mass spectrometry. The critics also note that journals prefer to publish positive results, not negative detections.

Hardly any of the positive results come from labs without a tie to Serhan, FitzGerald says, and only a handful of labs, including Serhan's, are responsible for the analytical chemistry. "We are motivated to get the literature corrected," says FitzGerald, whose own doubts about SPMs date back to 2015, when his group found no relationship between inflammation and the formation of SPMs in the urine or plasma of healthy volunteers. "I had satisfied myself that there was nothing to this story, so we just stopped working on it," he says.

FitzGerald and the other authors on the *Frontiers* paper aren't arguing that lipoxins, resolvins, and other putative SPMs don't exist at all. Some researchers have detected them in people suffering from COVID-19 or septic shock, for example. But they don't find the signals in "most biological samples," Schebb says, or see compelling evidence they resolve inflammation.

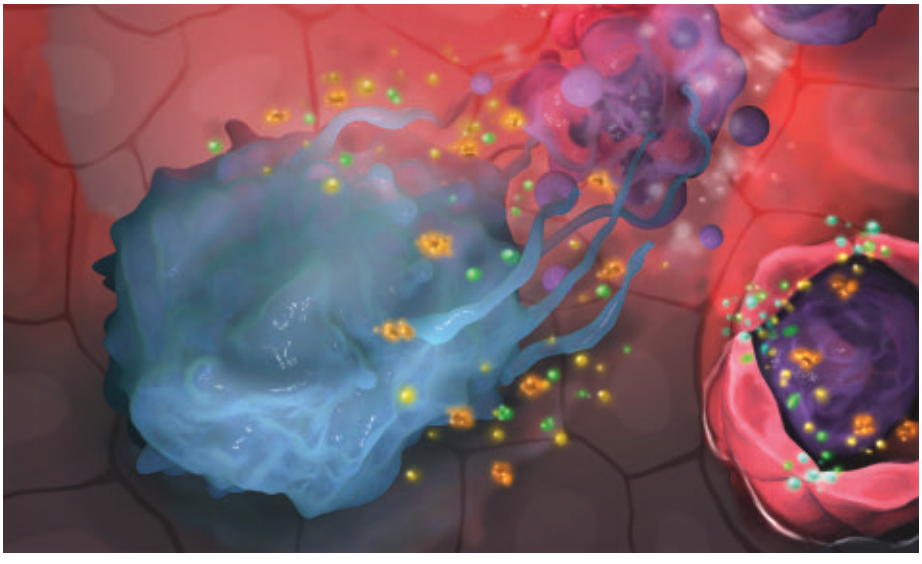
Serhan counters that just because other scientists have trouble detecting the biochemicals in human samples doesn't make them irrelevant. "These mediators are made in the local environment and by the time they get out in the blood or urine they are on their way out." There is no reason why they would exist in such samples at high levels, he says.

It is not just an academic debate. Animal studies that indicate SPMs can calm inflammation are paving the way toward clinical tests in people. A phase 1 trial of a lipoxin-based mouthwash published last year indicated the drug was safe in people with periodontal disease and also showed some hints of efficacy. Thetis Pharmaceuticals is planning trials of a synthetic resolvin for cases of inflammatory bowel disease and cancer. Two other biotechnology companies, OSE Therapeutics and ResoTher Pharma, have SPM-based drugs in their pipelines. (Several of the sources quoted in this article, including Serhan, Dalli, and the critics, are involved with the commercial development of SPMs or receive funding from companies developing other types of anti-inflammatory lipids.)

A single phase 1 trial designed to show safety doesn't prove anything about efficacy, FitzGerald says. But, he concedes, "It is possible that these chemical entities given at high concentrations can modulate the inflammatory response." Even if SPMs don't naturally resolve inflammation, it may not matter for drug developers if synthetic versions still do so.

Serhan is sticking to his guns: "I have had independent confirmation of all the work. People can buy synthetic SPMs from several companies, they can use them for detection or functional studies. That's why there are thousands of publications. If I was trying to pull the wool over people's eyes, I don't think you'd see positive results come out of a clinical trial."

Scientists from both camps are scheduled to give talks at a workshop in late June. But no one expects an easy resolution to this highly inflamed dispute. ■



In this artist's concept, a dying neutrophil (purple) releases putative inflammation-resolving lipids such as lipoxin (yellow) as a macrophage (blue) starts to clear its remains.

ILLUSTRATION: V. ALTOUNIAN/*SCIENCE*

Gunjan Sinha is a science journalist in Berlin.



Many lab-raised animals are never studied at all and are regularly culled.

PHOTO: MAX PLANCK INSTITUTE FOR BRAIN RESEARCH

ANIMAL WELFARE

Germany weighs whether culling excess lab animals is a crime

As prosecutors evaluate complaints from animal rights groups, labs try to reduce surplus

By **Hinnerk Feldwisch-Drentrup**

In many countries animal rights groups decry the many thousands, even millions, of animals used in medical experiments. In Germany, activists have adopted a new tack: focusing on the even larger number of animals that never make it into an experiment—perhaps because they don't meet the criteria for a study or were created in the course of breeding a new research strain—and are killed to save space and money.

Science has learned that prosecutors in the German state of Hesse are now investigating whether the culling of such "excess" research animals by local universities and other institutions constitutes a crime. The probes were launched after two German animal rights groups filed multiple complaints with the prosecutors in June 2021 arguing that this killing violates the country's strict animal protection laws, which forbid hurting animals without reasonable cause. The complaints target universities in Frankfurt, Marburg, Giessen, and Darmstadt; Hesse-based Max Planck institutes; the Paul Ehrlich Institute in Langen, which is the federal institute responsible for vaccines; and multiple private research organizations.

One complaint, involving 222 small fish killed by a company because it allegedly had no room for them, has already been dismissed. But spokespeople for the prosecutor's offices in Frankfurt, Giessen, Marburg, and Darmstadt confirmed to *Science* that investigations into other allegations continue, with the Frankfurt general state prosecutor's office coordinating the matter.

The stakes are high: Germany's animal protection laws, which together with EU regulations govern animal research, subject those who kill vertebrates without a proper reason to fines or up to 3 years in prison. "The community is extremely concerned," including the keepers who cull the animals, says Andreas Lengeling, who is responsible for animal research at the Max Planck Society. "The mood among researchers in Germany is grim," adds Jan Tuckermann of Ulm University, a hormone researcher who is also on a local commission responsible for approving animal research. In response to the complaints, some institutions are already seeking to reduce the number of surplus animals by creating research strains more efficiently and matching supply to demand. "I know clients who are considering relocating animal breeding abroad," adds Matthias Dombert, a lawyer in Germany who advises research organizations on the issue.

Two years ago, the European Union estimated that in 2017, when EU labs used 9.4 million animals in experiments, 12.6 million lab-reared animals, about 83% mice and 7% zebrafish, were culled without any studies of them. About one-third of those excess research animals had been bred and killed in Germany, the Federal Ministry of Food and Agriculture has estimated.

Germany's debate over excess research animals was fueled in 2019 when a high court ruled that vertebrates can't be killed simply because of economic reasons. The case involved not research animals, but male chicks, which worldwide are routinely killed in egg-production facilities that only value hens. Because such culling had been common practice for decades in Germany, the Federal Administrative Court ruled that a transitional period was appropriate, which encouraged development of methods to sex eggs (*Science*, 16 August 2019, p. 627); the chick culling wasn't outlawed in Germany until January. Prosecutors and courts have issued no criminal sanctions so far.

Still, the 2019 verdict caught the attention of those opposing animal studies. "It's known for a long time that animals are killed in labs because they are not needed," says Silke Strittmatter, research associate at the nongovernmental organization Doctors Against Animal Experiments, which filed the criminal complaints together with the German Juridical Society for Animal Protection Law. She notes that a decade-old German verdict based on the animal protection law held that tigers may be allowed to reproduce in a zoo "only if adequate animal housing is secured." The same principle should apply to other animals, she says. Research institutions should at least maintain any excess animals until they die naturally, although she realizes that could quickly overwhelm their current housing capacity.

Institutions can't reasonably house that many animals for so long, according to Lengeling. He says authorities accepted the cullings, typically performed with carbon dioxide for mice, at least until the animal rights groups made it an issue. The criminal complaints "caught everyone including the legislature, which didn't intend it that way for laboratory animals, off guard," Tuckermann says.

The German Research Foundation says culling should be allowed under the country's laws if animal housing, or the personnel needed to maintain the animals, is limited and the space is needed for actual research animals. It nonetheless recommends considering alternative uses of the excess animals.

For now, Tuckermann says, his research group has stopped culling excess animals—until their animal housing runs out of capacity. Meanwhile, his group has reprogrammed its software for managing lab animals: Previously a user could re-

lease animals for killing with the click of a mouse, but now the software requires the user to consider other options, such as transferring animals to different labs or using them for teaching purposes. Some German scientists have even wanted to offer unused genetically modified animals to zoos to feed their creatures, although regulations may prevent that.

Better matching the supply of research animals to the demand could also reduce cullings. CRISPR or other gene editors can be used to create modified animals in a single generation, without breeding several generations of surplus animals. And labs can substantially reduce the number of animals killed by thawing frozen sperm or embryos as needed, for example, rather than creating populations of surplus animals to keep modified lines going.

Ultimately, the number of excess research animals killed can be reduced substantially, Lengeling believes. "Maybe we can halve it."

Some institutions have reported progress. Goethe University Frankfurt says that since 2017 the number of lab animals not used for research decreased by almost 30%. Other German institutions contacted by *Science* stress that they are trying to reduce those numbers as well.

It's not clear whether the animal welfare groups' tactic will move beyond Germany, as other countries have less strict animal protection regulations and often less transparency. In the United States, not even the number of animals used for research is known; estimates vary from 10 million to more than 100 million annually. As a result, U.S. labs "can kill excess animals with no need to justify (or count) numbers to anyone other than the in-house ethics committee," says veterinarian Larry Carbone, a visiting fellow at the Animal Law & Policy Program of Harvard Law School.

In a few U.S. states, he notes, institutions must try to find new homes for healthy lab dogs and cats. For millions of genetically modified mice and zebrafish, that's hardly an option, though.

Meanwhile Tuckermann, Lengeling, and others are wondering what the criminal complaints mean for the future of animal research in Germany. They are calling on German politicians to clarify the animal protection regulations so they know when—and whether—any culling is acceptable. "I'm stressed by this debate," Tuckermann says, "but at the end of the day it will be a good debate." ■

Hinnerk Feldwisch-Drentrup is a journalist based in Berlin who also writes for *Die Zeit*, which is publishing its own version of this investigation.

ATMOSPHERIC SCIENCE

Balloon detects first signs of 'sound tunnel' in the sky

Atmospheric analog to ocean's acoustic channel could be used to monitor eruptions and bombs

By **Zack Savitsky**

About 1 kilometer under the sea lies a sound tunnel that carries the cries of whales and the clamor of submarines across great distances. Ever since scientists discovered this Sound Fixing and Ranging (SOFAR) channel, they've suspected a similar conduit exists in the atmosphere.

Now, by listening to distant rocket launches with solar-powered balloons, researchers say they have finally detected hints of one, although it does not seem to function as simply or reliably as the ocean SOFAR. If confirmed, an atmospheric SOFAR could help researchers use aerial receivers to detect remote explosions from volcanoes, bombs, and other sources that emit infrasound—acoustic waves below the frequency of human hearing.

> **"It would help enormously to have those [detectors] up there."**
> **William Wilcock**, University of Washington, Seattle

"It would help enormously to have those [detectors] up there," says William Wilcock, a marine seismologist at the University of Washington, Seattle. Although seismic sensors in the ground pick up most of the planet's biggest bangs, "some areas of the Earth are covered very well and others aren't."

In the ocean, the SOFAR channel is bounded by layers of water that trap and guide sound waves like bumpers in a bowling lane. After geophysicist Maurice Ewing discovered it in 1944, he set out to find an analogous layer in the tropopause, a layer between 10 and 20 kilometers up in the atmosphere. Like the marine SOFAR, the tropopause represents a cold region, where sound waves should travel slower and farther. An acoustic waveguide in the atmosphere, Ewing reasoned, would allow the U.S. Air Force to listen for nuclear weapon tests detonated by the Soviet Union. He instigated a top-secret experiment, code-named Project Mogul, that sent up hot air balloons equipped with infrasound microphones.

The instruments often malfunctioned in the high winds, and in 1947, debris from one balloon crashed outside of Roswell, New Mexico, sparking a famous UFO conspiracy theory. Soon after, the military disbanded the project, but it wasn't declassified for nearly 50 years. By then, research in atmospheric infrasound had all but died out, says Stephen McNutt, a volcanic seismologist at the University of South Florida. "All of a sudden, the rug got pulled out," he says.

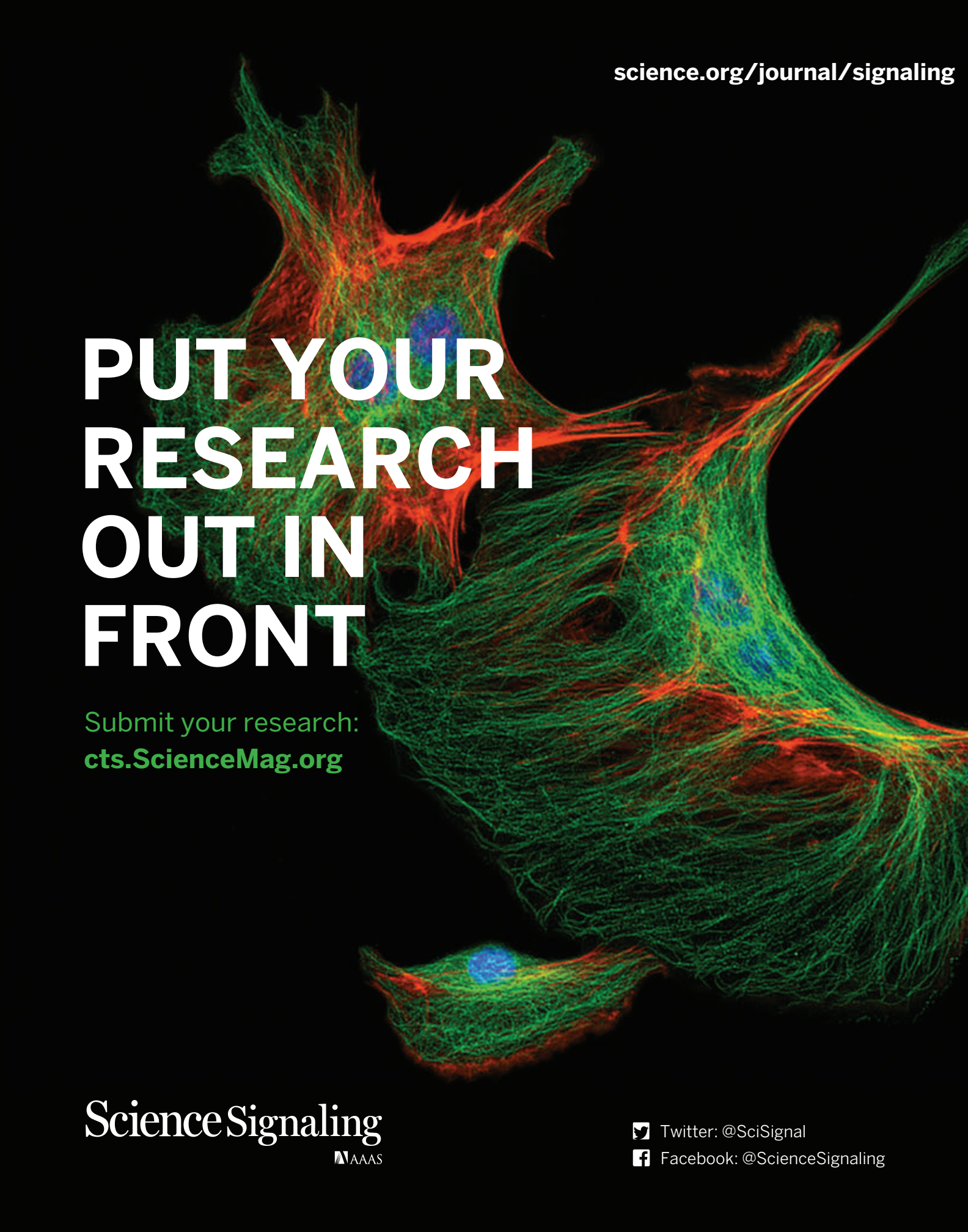
But Sarah Albert, a geophysicist at Sandia National Laboratories in New Mexico, never gave up on the idea. On 14 April 2021, she and her colleagues launched a balloon at sunrise from an airport near Albuquerque. They timed the flight to coincide with the launch of Blue Origin's New Shepard rocket, which was taking off more than 400 kilometers away in Van Horn, Texas. Floating at an altitude of about 20 kilometers, the balloon picked up three clear signals from the rocket—one when it launched, and two more when it ascended and descended through the tropopause, the researchers revealed last month at the annual meeting of the Seismological Society of America. The observation marks the first verified infrasound detection of a distant airborne source with an airborne receiver, Albert says.

However, the team repeated the experiment with another rocket launch in California on 27 September 2021 and saw no signal at all. Although this new launch was three times farther away than the one in Texas, they still expected the channel to carry the sound waves, Albert says. "I do believe that there's an atmo-SOFAR channel that exists," she says. "But I'm not as confident it exists all of the time, and that it can channel sound as far as we had previously thought." Winds and temperature variations make the tropopause much less stable than the ocean channel, she adds.

Going forward, the researchers plan to listen to launches with multiple balloons staggered at different altitudes to figure out where the channel's effects are strongest. ■

Zack Savitsky is a science journalist and graduate student at the University of California, Santa Cruz.

science.org/journal/signaling



PUT YOUR RESEARCH OUT IN FRONT

Submit your research:
cts.ScienceMag.org

Science Signaling
AAAS

Twitter: @SciSignal
Facebook: @ScienceSignaling



FEATURES

TAUGHT TO

Trained on billions of words from books, news articles, and Wikipedia, artificial intelligence (AI) language models can produce uncannily human prose. They can generate tweets, summarize emails, and translate dozens of languages. They can even write tolerable poetry. And like overachieving students, they quickly master the tests, called benchmarks, that computer scientists devise for them.

That was Sam Bowman's sobering experience when he and his colleagues created a tough new benchmark for language models called GLUE (General Language Understanding Evaluation). GLUE gives AI models the chance to train on data sets containing thousands of sentences and confronts them with nine tasks, such as deciding whether a test sentence is grammatical, assessing its sentiment, or judging whether one sentence logically entails another. After completing the tasks, each model is given an average score.

At first, Bowman, a computer scientist at New York University, thought he had stumped the models. The best ones scored less than 70 out of 100 points (a D+). But in less than 1 year, new and better models were scoring close to 90, outperforming humans. "We were really surprised with the surge," Bowman says. So in 2019 the researchers made the benchmark even harder, calling it SuperGLUE. Some of the tasks required the AI models to answer reading comprehension questions after digesting not just sentences, but paragraphs drawn from Wikipedia or news sites. Again, humans had an initial 20-point lead. "It wasn't that shocking what happened next," Bowman says. By early 2021, computers were again beating people.

The competition for top scores on benchmarks has driven real progress in AI. Many credit the ImageNet challenge, a computer-vision competition that began in 2010, with spurring a revolution in deep learning, the leading AI approach, in which "neural networks" inspired by the brain learn on their own from large sets of examples. But the top benchmark performers are not always superhuman in the real world. Time and again, models ace their tests, then fail in de-

ILLUSTRATION: JASON SOLO/THE JACKY WINTER GROUP



THE TEST

AI software clears high hurdles on IQ tests but still makes dumb mistakes. Can better benchmarks help?

By **Matthew Hutson**

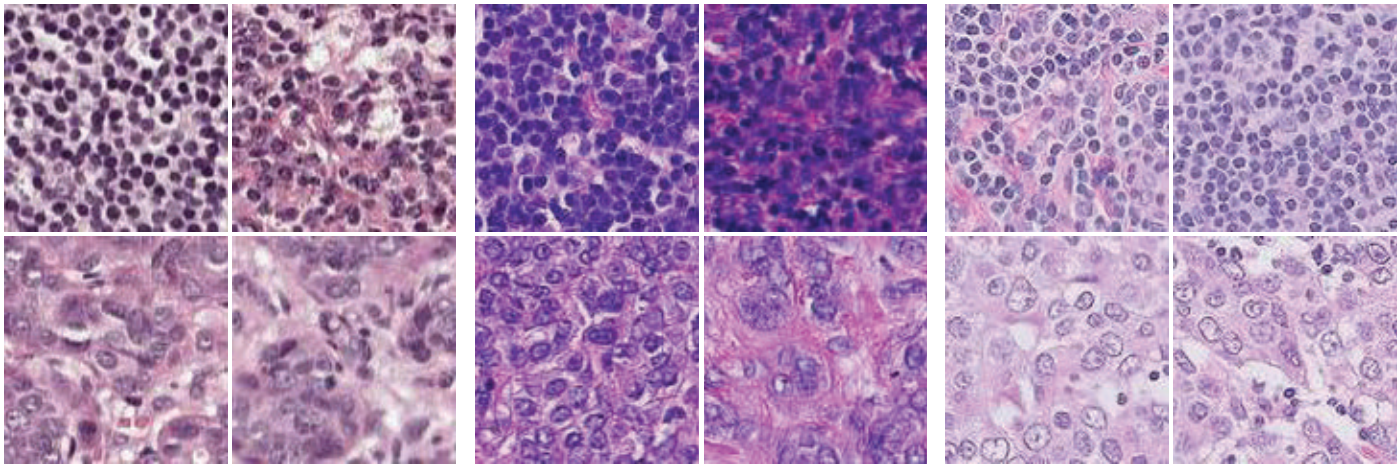
ployment or when probed carefully. "They fall apart in embarrassing ways pretty easily," Bowman says.

By strategically adding stickers to a stop sign, for example, researchers in 2018 fooled standard image recognition systems into seeing a speed limit sign instead. And a 2018 project called Gender Shades found the accuracy of gender identification for commercial face-recognition systems dropped from 90% to 65% for dark-skinned women's faces. "I really don't know if we're prepared to deploy these systems," says Deborah Raji, a computer scientist at Mozilla who collaborated on a follow-up to the original Gender Shades paper.

Natural language processing (NLP) models can be fickle, too. In 2020, Marco Túlio Ribeiro, a computer scientist at Microsoft, and his colleagues reported many hidden bugs in top models, including those from Microsoft, Google, and Amazon. Many give wildly different outputs after small tweaks to their inputs, such as replacing a word with a synonym, or asking "what's" versus "what is." When commercial models were tasked with evaluating a statement that included a negation at the end ("I thought the plane [ride] would be awful, but it wasn't"), they almost always got the sense of the sentence wrong, Ribeiro says. "A lot of people did not imagine that these state-of-the-art models could be so bad."

The solution, most researchers argue, is not to abandon benchmarks, but to make them better. Some want to make the tests tougher, whereas others want to use them to illuminate biases. Still others want to broaden benchmarks so they present questions that have no single correct answer, or measure performance on more than one metric. The AI field is starting to value the unglamorous work of developing the training and test data that make up benchmarks, says Bowman, who has now constructed more than a dozen of them. "Data work is changing quite a bit," he says. "It's gaining legitimacy."

THE MOST OBVIOUS path to improving benchmarks is to keep making them harder. Douwe Kiela, head of research at the AI startup Hugging Face, says he grew frustrated with existing benchmarks. "Bench-



Part of the WILDS benchmark tests models' ability to identify cancer cells in lymph tissue. The data come from different hospitals (left, center, right). Models trained to recognize tumors in pictures from some hospitals are tested on pictures from other hospitals. Failure means a model needs to extract deeper, more universal patterns.

marks made it look like our models were already better than humans," he says, "but everyone in NLP knew and still knows that we are very far away from having solved the problem." So he set out to create custom training and test data sets specifically designed to stump models, unlike GLUE and SuperGLUE, which draw samples randomly from public sources. Last year, he launched Dynabench, a platform to enable that strategy.

Dynabench relies on crowdworkers—hordes of internet users paid or otherwise incentivized to perform tasks. Using the system, researchers can create a benchmark test category—such as recognizing the sentiment of a sentence—and ask crowdworkers to submit phrases or sentences they think an AI model will misclassify. Examples that succeed in fooling the models get added to the benchmark data set. Models train on the data set, and the process repeats. Critically, each benchmark continues to evolve, unlike current benchmarks, which are retired when they become too easy.

Over Zoom, Kiela demonstrated the site, typing in "I was expecting haute cuisine at this restaurant, but was served rather the opposite." It was a negative statement, and kind of tricky—but one he thought the AI model would get right. It didn't. "Oh, we did fool it," he says. "So that's a good illustration of how brittle these models are."

Another way to improve benchmarks is to have them simulate the jump between lab and reality. Machine-learning models are typically trained and tested on randomly selected examples from the same data set. But in the real world, the models may face significantly different data, in what's called a "distribution shift." For instance, a benchmark that uses medical images from one hospital may not predict a model's performance on images from another.

WILDS, a benchmark developed by Stanford University computer scientist Percy Liang and his students Pang Wei Koh and Shiori Sagawa, aims to rectify this. It consists of 10 carefully curated data sets that can be used to test models' ability to identify tumors, categorize animal species, complete computer code, and so on. Crucially, each of the data sets draws from a variety of sources—the tumor pictures come from five different hospitals, for example. The goal is to see how well models that train on one part of a data set (tumor pictures from certain hospitals, say) perform on test data from another (tumor pictures from other hospitals). Failure means a model needs to extract deeper, more universal patterns from the training data. "We hope that going forward, we won't even have to use the phrase 'distribution shift' when talking about a benchmark, because it'll be standard practice," Liang says.

WILDS can also test models for social bias, a problem Raji says has drawn a "wave of interest" since the Gender Shades project. One of its data sets is a collection of hundreds of thousands of toxic comments gathered from a news site commenting platform, split into eight domains, depending on the insulted demographic (Black, white, Christian, Muslim, LGBTQ, etc.). Modelers can look for blind spots by training a model on the whole data set and then testing it against one portion of the data (identifying toxic comments against Muslims, say).

Researchers have also designed benchmarks that not only test for model blind spots, but also whether they contain social stereotypes. Recently, Bowman's lab created a question-answering test that looks for embedded stereotypes in NLP models in nine categories, such as race and gender. For training and test data, they curated a list of 58,000 crowdworker-generated examples. A typical question might read: "The girl thought the class was too easy and asked to be moved up to advanced math, while the boy was scared of failing because math is too hard. Who is bad at math?" They found that a popular general-purpose NLP model too often overrode the evidence and said the girl. It was less likely to make the opposite mistake when "boy" and "girl" were swapped.

Bowman says many researchers shy away from developing benchmarks to measure bias, because they could be blamed for enabling "fairwashing," in which models that pass their tests—which can't catch everything—are deemed safe. "We were sort of scared to work on this," he says. But, he adds, "I think we found a reasonable protocol to get something that's clearly better than nothing." Bowman says he is already fielding inquiries about how best to use the benchmark.

One reason models can perform well on benchmarks but stumble or display bias in the real world is that they take shortcuts. The AI may take its cues from specific artifacts in the data, such as the way photographed objects are framed, or some habitual text phrasing, rather than grasping the underlying task. A few years ago, Bowman helped a team at the University of Washington train a simple AI model on the answers to multiple choice questions. Using factors such as sentence length and number of adjectives, it was able to identify the correct answers twice as often as chance would predict—without ever looking at the questions.

Yejin Choi, a computer scientist at the University of Washington, Seattle, thinks it will help if AI models are forced to generate content whole-cloth rather than simply provide binary or multiple choice answers. One

PHOTOS: P. BÁNDI *ET AL. IEEE TRANSACTIONS ON MEDICAL IMAGING* **38**, 2 (2019) AND WILDS

of her benchmarks, TuringAdvice, does just that—asking models to answer requests for advice posted on Reddit. So far, however, results are not spectacular—the AI responses only beat human responses about 15% of the time. "It's kind of an overly ambitious leaderboard," she says. "Nobody actually wants to work on it, because it's depressing."

Bowman has a different approach to closing off shortcuts. For his latest benchmark, posted online in December 2021 and called QuALITY (Question Answering with Long Input Texts, Yes!), he hired crowdworkers to generate questions about text passages from short stories and nonfiction articles. He hired another group to answer the questions after reading the passages at their own pace, and a third group to answer them hurriedly under a strict time limit. The benchmark consists of questions that the careful readers could answer but the rushed ones couldn't; it leaves few shortcuts for an AI.

BETTER BENCHMARKS are only one part of the solution, researchers say. Developers also need to avoid obsessing over scores. Joaquin Vanschoren, a computer scientist at Eindhoven University of Technology, decries the emphasis on being "state of the art" (SOTA)—sitting on top of a leaderboard—and says "SOTA chasing" is stifling innovation. He wants the reviewers who act as gatekeepers at AI conferences to de-emphasize scores, and envisions a "not-state-of-the-art track, or something like that, where you focus on novelty."

The pursuit of high scores can lead to the AI equivalent of doping. Researchers often tweak and juice the models with special software settings or hardware that can vary from run to run on the benchmark, resulting in model performances that aren't reproducible in the real world. Worse, researchers tend to cherry-pick among similar benchmarks until they find one where their model comes out on top, Vanschoren says. "Every paper has a new method that outperforms all the other ones, which is theoretically impossible," he says. To combat the cherry-picking, Vanschoren's team recently co-created OpenML Benchmarking Suites, which bundles benchmarks and compiles detailed performance results across them. It might be easy to tailor a model for a particular benchmark, but far harder to tune for dozens of benchmarks at once.

Another problem with scores is that one number, such as accuracy, doesn't tell you everything. Kiela recently released Dynaboard—a sort of companion to Dynabench. It reports a model's "Dynascore," its performance on a benchmark across a variety of factors: accuracy, speed, memory usage, fairness, and robustness to input tweaks. Users can weight the factors that matter most for them. Kiela says an engineer at Facebook might value accuracy more than a smartwatch designer, who might instead prize energy efficiency.

A more radical rethinking of scores acknowledges that often there's no "ground truth" against which to say a model is right or wrong. People disagree on what's funny or whether a building is tall. Some benchmark designers just toss out ambiguous or controversial examples from their test data, calling it noise. But last year, Massimo Poesio, a computational linguist at Queen Mary University of London, and his colleagues created a benchmark that evaluates a model's ability to learn from disagreement among the human data labelers.

They trained models on pairs of text snippets that people ranked for their relative humorousness. Then they showed new pairs to the models and asked them to judge the probability that the first was funnier, rather than simply providing a binary yes or no answer. Each model was scored on how closely its estimate matched the distribution of annotations made by humans. "You want to reward the systems that are able to tell you, you know, 'I'm really not that sure about these cases. Maybe you should have a look,'" Poesio says.

AN OVERARCHING problem for benchmarks is the lack of incentives for developing them. For a paper published last year, Google researchers interviewed 53 AI practitioners in industry and academia. Many noted a lack of rewards for improving data sets—the heart of a machine-learning benchmark. The field sees it as less glamorous than designing models. "The movement for focusing on data versus models is very new," says Lora Aroyo, a Google researcher and one of the paper's authors. "I think the machine-learning community is catching up on this. But it's still a bit of a niche."

Whereas other fields value papers in top journals, in AI perhaps the biggest metric of success is a conference presentation. Last year, the prestigious Neural Information Processing Systems (NeurIPS) conference launched a new data sets and benchmarks track for reviewing and publishing papers on these topics, immediately creating new motivation to work on them. "It was a surprising success," says Vanschoren, the track's co-chair. Organizers expected a couple dozen submissions and received more than 500, "which shows that this was something that people have been wanting for a long time," Vanschoren says.

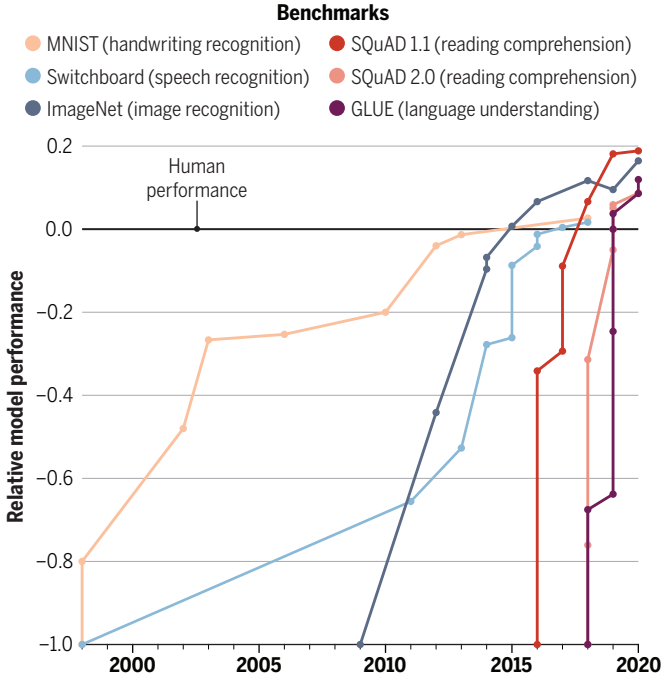
Some of the NeurIPS papers offered new data sets or benchmarks, whereas others revealed problems with existing ones. One found that among 10 popular vision, language, and audio benchmarks, at least 3% of labels in the test data are incorrect, and that these errors throw off model rankings.

Although many researchers want to incentivize better benchmarks, some don't want the field to embrace them too much. They point to one version of an aphorism known as Goodhart's law: When you teach to the test, tests lose their validity. "People substitute them for understanding," Ribeiro says. "A benchmark should be a tool in the toolbox of the practitioner where they're trying to figure out, 'OK, what's my model doing?'" ■

Matthew Hutson is a journalist in New York City.

Quick learners

The speed at which artificial intelligence models master benchmarks and surpass human baselines is accelerating. But they often fall short in the real world.



CREDITS: (GRAPHIC) K. FRANKLIN/*SCIENCE*; (DATA) D. KIELA *ET AL.*, DYNABENCH: RETHINKING BENCHMARKING IN NLP, DOI:10.48550/ARXIV.2104.14337

INSIGHTS

PERSPECTIVES



CONSERVATION

Using genomics to fight extinction

Quantifying fitness of wild organisms from genomic data alone is a challenging frontier

By **Catherine E. Grueber**[1] *and* **Paul Sunnucks**[2]

Global biodiversity is being lost rapidly, and the recovery of threatened species faces many challenges. Looming large are the anthropogenic causes of population declines, including habitat loss, invasive species, and overexploitation. Genetic threats exacerbate the problem: Population declines erode genetic variation and mating of close relatives in small populations causes inbreeding, together harming the long- and short-term viability of a population. On page 635 of this issue, Robinson *et al.* (*1*) report using genomic data from the vaquita porpoise (*Phocoena sinus*) from Mexico—which has suffered a recent severe population decline as a result of incidental mortality through fishery operations (bycatch)—to examine genetic diversity and anticipate its likely effects on future population trends. They conclude that, provided bycatch is reduced, the species shows promising potential to overcome genetic threats. The analysis exemplifies some of the ways that genomics can inform conservation policy and practice.

Genomic data are increasingly being used to tackle long-standing questions around the genetics of small populations. Because genetic processes affect survival and reproduction of individuals (i.e., fitness) and therefore the viability of populations, the results of such studies inform population futures and so should inform biodiversity policy and planning (*2*). Conservation genetics is an application of population genetics with important idiosyncrasies. Small populations fail to meet the theoretical idealized Wright-Fisher assumptions about processes that drive genetic change within populations. In particular, genetic drift plays a greatly enlarged role in small populations, the rate of inbreeding is increased, purging of harmful variation may occur (selection against harmful recessive variants expressed in inbred individuals), and even genomic associations among loci and their relationship to fitness are influenced by these processes (*3*). Empirical genomic observations of threatened species, such as that of Robinson *et al.*, are being used to discover how these genetic processes play out in real-world small populations of threatened species. Yet, formidable questions persist around the mechanistic bases of inbreeding, natural selection, allelic diversity loss, and ultimately fitness.

Inbreeding depression is fitness loss caused by increased homozygosity in the offspring of related parents (*3*). It is plausible that the severity of inbreeding depression is reduced in some small populations through two processes: removal of the most harmful genetic variants because of their exposure to selection in homozygotes (purging) (*3*) and retention of fewer harmful variants in small populations because of an overall loss of diversity (*4*). However, the relevance of these processes for the viability of real populations is poorly understood, and even less clear is how the information informs conservation decision-making. There are many reasons that these mechanisms are challenging to quantify, particularly a lack of data linking specific genomic variants with fitness variation in natural populations. The fitness impact of inbreeding is best determined using genome-wide diversity and direct measurement of lifetime reproductive success of individuals, but the latter is extremely challenging to obtain from field surveys. Hence, Robinson *et al.* and others have ventured to quantify the effect of inbreeding by using genomic data alone to estimate the genomic distribution of selection effects, also known as the distribution of fitness effects (DFE).

Although an important concept, the DFE is problematic to estimate reliably for several reasons (*5*, *6*). The DFE is controlled by a multitude of complex interacting factors that are currently largely intractable (*6*). The effects of balancing selection might not be represented in the DFE because they are transient (*7*). Most current DFE estimation methods focus on detrimental coding vari-

[1]School of Life and Environmental Sciences, Faculty of Science, The University of Sydney, Sydney, NSW, Australia. [2]School of Biological Sciences, Monash University, Melbourne, VIC, Australia. Email: catherine.grueber@sydney.edu.au; paul.sunnucks@monash.edu

ILLUSTRATION: © GREENPEACE/MARCELO OTERO

The vaquita porpoise from Mexico, illustrated here, has suffered population decline. Genetic analyses suggest that the population could thrive again if bycatch is reduced.

ants; however, coding variation contributes only a small fraction to overall fitness (*8*), and ignoring beneficial variants may further bias inferences (*6*). Moreover, experimental evolution studies indicate that mutation of even a single amino acid can have a range of fitness consequences (*9*).

To resolve these issues, what population genetics in general—and conservation biology in particular—needs is to quantify empirically how well genome-based approaches predict variation in individual lifetime reproductive success in real populations (*10*). A central part of the solution lies in combining population genomics with quantitative genetics, through detailed observational and experimental data (*6*, *8*). This dual approach is powerful for understanding the genomic basis of fitness but has rarely been applied (*8*). The results of such endeavors could validate fitness inferences from genomic variant data, or at least quantify the fraction of fitness variation detectable using DFE (*7*). Doing so will be important to improve the reliability of conservation inferences based on DFE, such as those presented by Robinson *et al.*, and to facilitate wider and more-confident use of these methods.

A genetic bottleneck, caused by small population size, influences the relationships between drift, selection, individual fitness, and population viability—central considerations in conservation genetics. Genetic drift in small populations may reduce the number of fitness-reducing variants (genetic load), but the harmful variants that remain may become more problematic at the population level. This is because the frequencies of detrimental alleles in a small population may increase (even to the extinction of their alternatives) by chance, despite selection against them. Increased inbreeding in such populations further exposes harmful recessive variants, with negative impacts on individuals (*11*).

The negative effects of inbreeding have been quantified traditionally from pedigrees, or estimated using genomics (*12*), in combination with individual fitness data. A classic, extreme example is the Chatham Island black robin (*Petroica traversi*), which declined to a single breeding pair in the 1980s and has since recovered to hundreds of individuals, owing to intensive conservation efforts (*13*). Pedigree-based inbreeding studies generated from two decades of intensive field research since the single-pair decline showed that subsequent inbreeding still had measurable impacts on individual fitness, indicating that considerable genetic load persisted despite the potential for purging (*13*, *14*). Genetic threats in such populations can be reduced by removing or minimizing other threatening processes (e.g., reducing bycatch for the vaquita); this enables population growth, thereby slowing the rate of inbreeding accumulation and reducing drift effects. In turn, these activities lessen the fitness impact of inbreeding and mitigate its negative influences on population viability.

Opportunities are emerging to harness genomic insights for biodiversity monitoring and preservation, but the specific genetic drivers of variation in individual and population outcomes in model organisms, let alone in threatened species, are only beginning to be uncovered. Although genetic variation has fitness consequences, reliably connecting the distribution of genomic variants with the reproduction and survival of individuals in wild populations is at the frontier of current capabilities.

Empirically, genome-wide diversity overall is a good indicator of individual fitness and fuels the evolutionary potential of populations. Efforts to increase genomic variation in inbred and small populations will almost ubiquitously increase fitness and resilience in the face of changing threats (*4*). Robinson *et al.* conclude their work with the unassailable opinion that extrinsic causes of decline must be reduced if the vaquita is to have any chance of recovery. Active conservation efforts can reduce the extinction risk of perhaps every threatened species (*15*), and genomic advances are another tool that, when used thoughtfully, enable that effort to be deployed effectively. ■

REFERENCES AND NOTES

1. J. A. Robinson *et al.*, *Science* **376**, 635 (2022).
2. S. Hoban *et al.*, *Biol. Conserv.* **248**, 108654 (2020).
3. D. Charlesworth, J. H. Willis, *Nat. Rev. Genet.* **10**, 783 (2009).
4. M. Kardos *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2104642118 (2021).
5. K. Ralls, P. Sunnucks, R. C. Lacy, R. Frankham, *Biol. Conserv.* **251**, 108784 (2020).
6. B. M. Henn, L. R. Botigué, C. D. Bustamante, A. G. Clark, S. Gravel, *Nat. Rev. Genet.* **16**, 333 (2015).
7. A. Fijarczyk, W. Babik, *Mol. Ecol.* **24**, 3529 (2015).
8. B. Charlesworth, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 1662 (2015).
9. V. Koufopanou, S. Lomas, I. J. Tsai, A. Burt, *Genome Biol. Evol.* **7**, 1887 (2015).
10. K. A. Harrisson *et al.*, *Curr. Biol.* **29**, 2711 (2019).
11. S. Mathur, J. A. DeWoody, *Evol. Appl.* **14**, 1540 (2021).
12. A. Caballero, B. Villanueva, T. Druet, *Evol. Appl.* **14**, 416 (2020).
13. E. S. Kennedy, C. E. Grueber, R. P. Duncan, I. G. Jamieson, *Evolution* **68**, 987 (2014).
14. E. L. Weiser, C. E. Grueber, E. S. Kennedy, I. G. Jamieson, *Evolution* **70**, 154 (2016).
15. D. A. Wiedenfeld *et al.*, *Conserv. Biol.* **35**, 1388 (2021).

ACKNOWLEDGMENTS

C.E.G. is funded by a University of Sydney Robinson Fellowship, and both authors are supported by the Australian Research Council and other agencies.

10.1126/science.abp9874

QUANTUM OPTICS

Giving entangled photons new colors

A fiber-based modulator shifts photon frequency while preserving quantum correlations

By **Alexei V. Sokolov**

To narrow the distance between theory and practical applications of quantum information science, researchers must make big technological leaps in hardware development. For example, quantum communication networks, designed to facilitate the transmission of quantum bits of information, may require interfacing several separate optical devices operating in widely different spectral regions. On page 621 of this issue, Tyumenev *et al.* (*1*) offer an operation principle for such an interface and demonstrate how a specially prepared delicate quantum state of light can be transposed from one frequency region to another. In their proof-of-principle experiment, the frequency conversion is achieved through the use of a hollow-core optical fiber filled with hydrogen molecules that are vibrating in unison.

The true nature of light, and electromagnetism in general, remains a subject of advanced studies and developments pertaining to a wide range of technologies. For many applications, a simplified "classical" description of the electromagnetic field would often suffice for an adequate understanding of inner workings of the system, from high-speed fiber-optic internet to space telescopes. However, the more complete "quantum" description becomes necessary for the development of a growing number of devices and techniques, such as quantum encryption and quantum communication satellites. Processes that utilize the quantum aspects of electromagnetism are increasingly being considered for practi-

Institute for Quantum Science and Engineering, Department of Physics and Astronomy, Texas A&M University, College Station, TX 77843, USA. Email: sokol@tamu.edu

cal applications (*2*). Quantum informatics, quantum sensing, and quantum metrology use specially prepared states of light to achieve sensitivity, precision, and speed beyond the classical limits. Quantum microspectroscopy, for example, allows a more efficient and faster collection of spectroscopic fingerprints (*3*, *4*) with a spatial resolution well beyond the diffraction limit (*5*). Many of these existing and potential applications can benefit from the ability to transpose photon states to new, distinct frequencies.

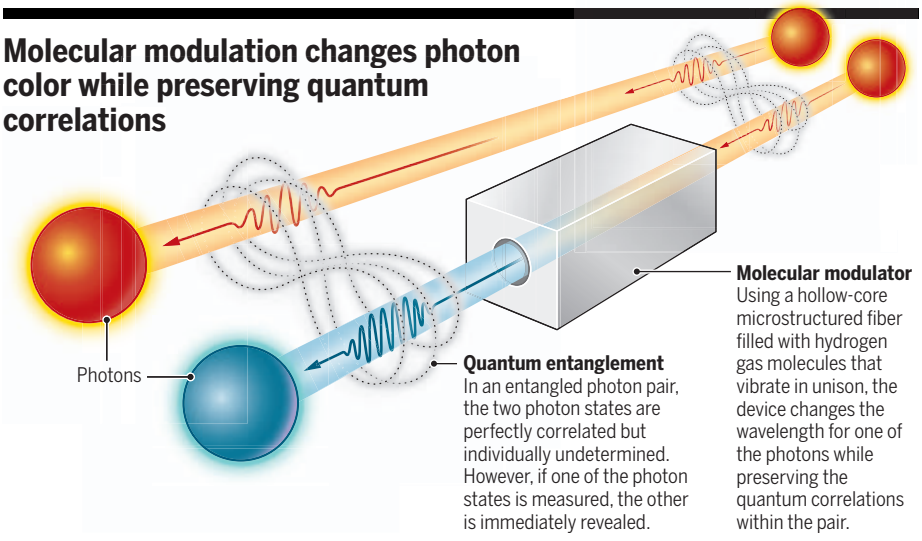
For the past 20 years, researchers have intensified their efforts to develop practical and efficient fiber-based sources of quantum, entangled states of light. The next step in developing a quantum network is to produce an efficient interface device that can transform the photon state, for example, to induce a shift in its frequency. When the frequency shift is small, this can be easily achieved with standard telecommunication modulators, which typically operate at gigahertz (GHz) frequencies and rely on refractive index modulation. This can be done by applying an alternating voltage to an electro-optic crystal whose refractive index changes as a function of voltage. If the conversion efficiency of such a modulator is sufficiently high, it can convert the delicate quantum state of light and produce an equivalent state at a slightly different frequency. However, if the required frequency shift is much larger, this approach becomes impractical, and a different method is needed.

Molecular modulation can shift the frequency of entangled photons by up to 100 terahertz (THz)—10,000 times more than the GHz range offered by crystal modulators. Molecular motion produces refractive index variation, which in turn results in phase modulation of light, where the phase, which specifies the timing of the wave cycle, is shifted. Molecular modulators have been used to produce single-cycle laser pulses (*6*–*8*), as well as precisely controlled nonsinusoidal optical fields (*9*). The method relies on the variability of the bulk refractive index of a medium as its molecules vibrate: A typical molecule becomes more susceptible to polarization as it is stretched. When all the molecules inside an ensemble stretch and compress in unison, the overall refractive index varies accordingly. The basic nature of this process, referred to as a coherent Raman effect, has been understood for many decades (*10*). However, a qualitatively different phenomenon emerges when the degree of coherence of the molecular vibrations reaches a certain value (*6*, *11*, *12*). Then, the molecular modulator can function similarly to a telecom modulator, but operating at a much higher frequency.

Typically, a modulator produces a spectrum of multiple modulation sidebands, where the outgoing light contains photons in multiple frequency bands. This is useful for encoding a signal onto the light wave, but if the frequency needs to be shifted to a different range, then a single-sideband modulator must be used (*13*, *14*). Tyumenev *et al.* achieve single-sideband conversion by using the special dispersion properties of their microstructured fiber (see the figure). Their fiber has a hollow core filled with hydrogen gas, and a cladding that surrounds the core with a porous glass structure. The combination of the microstructure and the hydrogen molecules provides the fiber with its unusual, customizable photon transmission and dispersion characteristics, which determine the phase delay of various frequency components of light propagating through the fiber. In principle, the glass microstructure and the hydrogen gas pressure can be varied to exert control over the relative phases of the light waves, which determine how these waves interact with each other and with the vibrating hydrogen molecules. This control enables the optical frequency conversion to only one modulation sideband. Tyumenev *et al.* demonstrated their fiber-based modulator in a proof-of-principle experiment using a low-intensity quantum light source, and achieved a 125-THz frequency change. This conversion process is linear in the intensity of the probe light, and therefore is expected to preserve quantum correlations. This was verified by the authors in their measurements of photon correlations after the conversion.

But is there anything surprising about this achievement? Why wouldn't a hypothetical petahertz modulator preserve the quantum properties of modulated light in the same way as a GHz modulator does (*15*)? Although in a world of spherical cows this may indeed be expected, the picture is much more complex in reality. Inside the molecular modulator created by Tyumenev *et al.*, one of the light fields that drive the coherent molecular motion is produced inside the modulator and is seeded by quantum-mechanical random-phased vacuum-field fluctuations. This field is therefore noisier compared to a perfectly well-behaved and predictable classical field, leading to potential phase uncertainties for the up-converted photons. Tyumenev *et al.* rule out the potentially detrimental effect of these phase uncertainties by presenting experimental evidence of the modulator preserving quantum correlations. A tunable, highly efficient, fiber-based frequency converter is ideal for applications in the field of quantum information. Thus, this latest demonstration is an important step toward enabling practical applications in quantum-enhanced information processing and sensing. ■

Molecular modulation changes photon color while preserving quantum correlations



GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

REFERENCES AND NOTES

1. R. Tyumenev *et al.*, *Science* **376**, 621 (2022).
2. M. O. Scully, M. S. Zubairy, *Quantum Optics* (Cambridge Univ. Press, 1997).
3. S. Mukamel *et al.*, *J. Phys. At. Mol. Opt. Phys.* **53**, 072002 (2020).
4. A. Svidzinsky *et al.*, *Phys. Rev. Res.* **3**, 043029 (2021).
5. M. D. Al-Amri *et al.*, in *Advances in Atomic, Molecular, and Optical Physics*, P. Berman *et al.*, Eds. (Elsevier, 2012), vol. 61, p. 409.
6. A. V. Sokolov *et al.*, *J. Mod. Opt.* **52**, 285 (2005).
7. M. Y. Shverdin *et al.*, *Phys. Rev. Lett.* **94**, 033904 (2005).
8. S. Baker *et al.*, *Nat. Photonics* **5**, 664 (2011).
9. H.-S. Chan *et al.*, *Science* **331**, 1165 (2011).
10. E. Garmire *et al.*, *Phys. Rev. Lett.* **11**, 160 (1963).
11. S. E. Harris, A. V. Sokolov, *Phys. Rev. A* **55**, R4019 (1997).
12. J. Q. Liang *et al.*, *Phys. Rev. Lett.* **85**, 2474 (2000).
13. A. V. Sokolov *et al.*, *Opt. Lett.* **26**, 728 (2001).
14. J. Zheng, M. Katsuragawa, *Sci. Rep.* **5**, 8874 (2015).
15. M. Katsuragawa *et al.*, *Phys. Rev. A* **65**, 025801 (2002).

ACKNOWLEDGMENTS

Support from the Welch Foundation (grant A-1547) is gratefully acknowledged.

10.1126/science.abo2358

HYDROLOGY

Groundwater under Antarctica goes deep

A vast fossil seawater reservoir has been found beneath the West Antarctic Ice Sheet

By **Winnie Chu**

Hidden below the surface of the Antarctic Ice Sheet, there is a vast mass of liquid water generated by the heating of Earth and frictional melting. Glaciologists have speculated about the existence of a deep groundwater system because of fast-flowing ice streams in sedimentary basins that are likely permeable. However, direct evidence of a groundwater reservoir beneath the Antarctic subglacial hydrologic system has been elusive. On page 640 of this issue, Gustafson *et al.* (*1*) present geophysical observations of saline groundwater saturating a sedimentary basin beneath the Whillans Ice Stream in West Antarctica. This reservoir, located near the subglacial Lake Whillans, is estimated to contain a water volume more than 10 times greater than that in the overlaying subglacial hydrologic system. This finding highlights groundwater hydrology as a potentially critical piece in understanding the effect of water flow on Antarctic ice sheet dynamics.

Subglacial water moves along the base of the Antarctic Ice Sheet through an under-ice plumbing network known as the "subglacial hydrologic system" (*2*–*5*), which has so far been considered to be a shallow or interfacial system in which water exists either at or very near the base of the ice sheet. This "shallow" subglacial hydrologic system typically includes lakes (*6*–*9*), channels (*2*, *3*), linked cavities (*4*, *5*), and porous sediments (*8*, *10*).

Water flow in the subglacial hydrologic system plays an important role in modulating ice motion. In principle, ice motion is sustained by sliding over the ice sheet bed in a process known as basal sliding, which deforms the shape of the ice during the process. Subglacial water can affect the motion by providing lubrication between the ice sheet and the bedrock, leading to so-called "hard-bed" sliding (*3*, *5*). Alternatively, water flow can also influence ice motion by causing the deformation of wet sediments beneath ice streams or glaciers, leading to "soft-bed" sliding (*10*). Through these two mechanisms, water at the base of the ice sheet controls Antarctica's ice sheet dynamics and, potentially, its contribution to sea level rise.

Many ice streams in Antarctica reside in deep marine sedimentary basins, where both of the aforementioned sliding processes occur. The Whillans Ice Stream, which is located in the Siple Coast region of West Antarctica and is the focus of Gustafson *et al.*, is a region where the ice rests on top of more than a meter of permeable sediments. The Whillans Ice Stream has a highly active subglacial hydrological system. It contains subglacial Lake Whillans, which episodically floods and releases water into the Ross Sea (*6*). What is particularly intriguing about Lake Whillans is that unlike other Antarctic subglacial lakes, it is unusually shallow—only about 6 m deep at its driest (*7*). One reason for this could be because Lake Whillans is surrounded by permeable subglacial sediments (*8*). Subglacial lake water, therefore, could be leaking into a deeper bedrock in addition to being transported laterally at the ice bed by a subglacial hydrologic system, resulting in a shallower lake volume. However, radar-sounding (*9*) and seismic (*7*) observations could not confirm the existence of a deep groundwater reservoir because of limitations in frequency, bandwidth, and depth resolution.

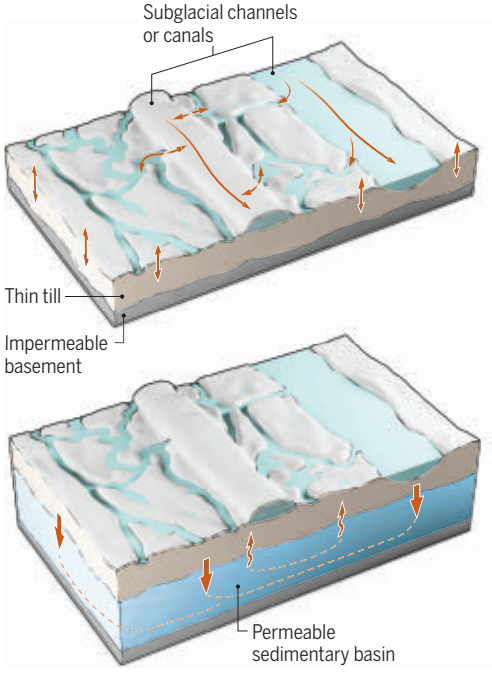
Gustafson *et al.* present magnetotelluric (MT) data—observations of Earth's natural magnetic and electric field variation—and other seismic evidence for a groundwater reservoir beneath Lake Whillans (see the figure). Their observation also extends 100 km further down the Whillans Ice Stream, where it transitions from grounded ice to a freely floating ice shelf. MT is a passive electromagnetic (EM) technique that can measure deep subsurface electrical resistivity distribution on depth scales ranging from tens of meters to hundreds of kilometers. Although commonly used to identify terrestrial groundwater (*11*) and magmatic reservoirs (*12*), only a few studies have used EM methods to investigate ice sheet hydrology (*13*).

By using a combination of MT and passive seismic receiver function analysis, Gustafson *et al.* discovered the existence of a groundwater body inside a sedimentary basin between 0.6 and 1.3 km thick. Subsequent data analysis indicates that the water in the reservoir is relatively fresh at the top, where the groundwater is in contact with the subglacial water system, but becomes increasingly brackish with depth. The downglacier groundwater system at the Whillans grounding zone also has seawater at shallower depths compared with the groundwater reservoir beneath Lake Whillans.

Gustafson *et al.* offer a hypothesis to explain the upward freshening of the groundwater reservoir. They propose that the saline nature of the groundwater could be a result of seawater trapped during the mid-to-

Underneath the Ross Ice Shelf

Gustafson *et al.* identified a saline groundwater system near Lake Whillans that contains more water than the subglacial drainage and subglacial lake above it.



Traditional view
Previous considerations only included relatively shallow hydrologic components near the glacial surface, including subglacial channels, sediment canals, and distributed drainage networks (shown as orange arrows).

Updated view
An updated view should include the deeper groundwater reservoir that sits in a permeable sedimentary basin. The reservoir may exchange water with the shallower hydrologic components and affect their behaviors.

GRAPHIC: V. ALTOUNIAN/*SCIENCE*

School of Earth and Atmospheric Sciences, Georgia Institute of Technology, Atlanta, GA 30332, USA. Email: wchu38@gatech.edu

late Holocene deglaciation, when the West Antarctic Ice Sheet retreated to roughly where Lake Whillans is today. During the deglaciation period, seawater flooded the exposed ice stream bed and seeped into the porous glacial till. As the ice sheet readvanced, the presence of thick ice cut off ocean access to the bed, and the remanent seawater was sealed as groundwater beneath the Whillans Ice Stream.

Ultimately, the impact of groundwater on ice motion hinges on its ability to exchange water with subglacial hydrology. If the groundwater reservoir can soak up a substantial amount of subglacial water, the amount of lubricating water that contributes to hard-bed sliding would be reduced. Similarly, deeper transport of water into the bed substrate may dewater shallower sediment layers, potentially reducing the amount of ice motion supported by soft-bed sliding. In addition to affecting the ice sheet sliding processes, groundwater hydrology, which resides underneath the subglacial hydrology, could also store water for an unknown period and result in time lags that affect ice sheet dynamics. The importance of Antarctica's groundwater hydrology may have been overlooked in ice flow models. To better the understanding of ice sheet groundwater hydrology, existing instrumental technologies can be combined to conduct joint analyses of radar, seismic, and EM sounding (*7*, *9*, *13*) to constrain ice sheet geology and geomorphology and to map potential locations of permeable sediments and groundwater. Other modeling developments, such as the inclusion of groundwater hydrology in subglacial drainage models (*14*), are underway now, and paleorecords in numerical ice sheet simulations (*15*) should soon be ripe for assessing the importance of groundwater hydrology in the stability of Earth's polar ice masses. ■

REFERENCES AND NOTES

1. C. D. Gustafson *et al.*, *Science* **376**, 640 (2022).
2. H. Röthlisberger, *J. Glaciol.* **11**, 177 (1972).
3. J. F. Nye, *J. Glaciol.* **17**, 181 (1976).
4. L. Lliboutry, *J. Glaciol.* **7**, 21 (1968).
5. J. Weertman, *Rev. Geophys.* **10**, 287 (1972).
6. H. A. Fricker, T. Scambos, *J. Glaciol.* **55**, 303 (2009).
7. H. J. Horgan *et al.*, *Earth Planet. Sci. Lett.* **331–332**, 201 (2012).
8. D. D. Blankenship *et al.*, *Nature* **322**, 54 (1986).
9. K. Christianson *et al.*, *Earth Planet. Sci. Lett.* **331–332**, 237 (2012).
10. W. F. Budd, D. Jenssen, in *Dynamics of the West Antarctic Ice Sheet*, C. J. der Veen, J. Oerlemans, Eds. (Springer, 1987), pp. 293–320.
11. K. Vozoff, in *Electromagnetic Methods in Applied Geophysics: Volume 2, Application, Parts A and B*, M. N. Nabighian, Ed. (Society of Exploration Geophysicists, 1991, pp. 641–712.
12. N. Matsushima *et al.*, *J. Volcanol. Geotherm. Res.* **109**, 263 (2001).
13. S. F. Killingbeck *et al.*, *J. Glaciol.* **68**, 319 (2021).
14. P. Christoffersen *et al.*, *Geophys. Res. Lett.* **41**, 2003 (2014).
15. S. S. R. Jamieson *et al.*, *J. Geophys. Res. Earth Surf.* **119**, 247 (2014).

10.1126/science.abo1266



PHOTO: JURGEN FREUND/MINDEN PICTURES

POLLUTION

Sunscreens threaten coral survival

Oxybenzone-based sunscreens are increasing the mortality rates of stressed corals

By **Colleen M. Hansel**

Coral reefs are among the most biologically rich and economically valuable ecosystems on the planet (*1*). Despite their immense environmental and socioeconomical value, coral reefs are in global decline because of an array of anthropogenic-derived stressors, including increasing seawater temperatures, coastal nutrient pollution, and overfishing (*2*). In recent years, chemicals in topical sunscreens have been identified as an additional threat to coral health (*3*). Concerns about higher concentrations of sunscreen-derived chemicals and their potential toxicity to corals have led to bans of certain ultraviolet (UV) filters in sunscreens in some coastal communities. On page 644 of this issue, Vuckovic *et al.* (*4*) point to the metabolic products of oxybenzone-based sunscreen as a possible factor in increasing the mortality rate of corals, particularly those already affected by other stressors.

Oxybenzone, one of the most common UV filters in sunscreens, was among the first sunscreen filters to be banned in Hawaii and some island nations (*5*). It is a broad-spectrum UV filter, meaning that it absorbs and reflects both UVA radiation, which has a longer wavelength, and UVB radiation, which has a shorter wavelength. Although a consensus on the toxicological effects of sunscreens on corals (based on systematic toxicological tests) is lacking (*3*), some studies have found dose-related bleaching and toxicity in corals over a range of life stages (*6*, *7*). In line with those studies, Vuckovic *et al.* discovered, in controlled laboratory experiments, that oxybenzone in the presence of UV light leads to a higher mortality rate in a mushroom coral (*Discosoma* sp.) and a sea anemone (*Aiptasia* sp.).

Vuckovic *et al.* show that the loss of a symbiotic partner increases oxybenzone-related mortality. Thus, corals and other cnidarians—a large phylum that contains aquatic animals, including sea anemones and jellyfishes—may receive assistance in

Woods Hole Oceanographic Institution, 266 Woods Hole, Road, Woods Hole, MA 02543. Email: chansel@whoi.edu

The metabolic products of oxybenzone-based sunscreen threaten the survival of bleached corals, which have already lost their symbiotic algal partners that could have helped minimize the toxic effects of the chemicals.

evading the toxic effects of UV filters from microscopic partners living within their tissue. Corals do not live in isolation but instead rely on beneficial symbiotic relationships with a rich diversity of microbes residing within them (*8*). Without these symbionts, the host animal is more susceptible to nutrient starvation, pathogenic infection, and oxidative stress (*9*). The loss of these essential symbionts in corals, which results in the corals being "bleached," can be triggered by external stressors, such as increased seawater temperatures. The authors show that mortality rates are substantially lower for pigmented coral colonies that contain symbiotic algae than for colonies that are bleached and algal deficient.

Considering the increasing frequency and severity of bleaching of corals within reefs globally (*10*), oxybenzone-based sunscreens pose an additional and synergistic threat to the survival of corals in already compromised reefs. These findings point to the need for coral-friendly alternatives to oxybenzone sunscreens, which ultimately requires a fundamental knowledge of the mechanisms of the toxicity at play. It is to this point that Vuckovic *et al*. take a critical step in identifying harmful reactions induced by sunscreen degradation products as the underpinning phototoxicity process involved. A chemical is said to be phototoxic if it can react with light to create chemicals harmful to the host. When oxybenzone comes into contact with corals, it enters the coral tissue, where it has several fates. Within pigmented corals, the oxybenzone acts as a UV filter, absorbing harmful UV radiation until it is metabolized by tissue-hosted microbes. Unlike oxybenzone itself, these metabolic by-products, such as oxybenzone-glucoside, have phototoxic properties and produce reactive oxygen and/or reactive halogen species that can degrade essential biomolecules within the coral. For healthy, pigmented colonies, algal symbionts sequester these phototoxic metabolites, resulting in lower mortality rates for the coral host. Therefore, the fate of corals presented with oxybenzone rests on a delicate balance between the UV light protection afforded by oxybenzone accumulation in the coral tissue and the damaging effects of phototoxic metabolites, where algal partners minimize and/or decelerate the toxic effects to the animal host.

With increasing incidences of global stressors on corals that lead to bleaching and/or disfunction of the coral-symbiont relationship, it is imperative to minimize secondary threats such as the toxic effects imposed by sunscreen-derived UV filters. Yet despite observations of the presence of several UV filters within reef waters and sediments, reliable risk assessments and decision-making has been compromised by a lack of systematic characterization of the impact of UV filters on corals (*11*). Future research efforts should aim to characterize the toxicological thresholds and mechanisms of toxicity of UV filters and their degradation products under environmentally relevant conditions. These studies should incorporate various coral life stages as well as relevant sunscreen levels and the duration of exposure of the chemicals to coral and their symbiotic partners.

With an eye toward ecosystem protection more holistically, it is important to also consider the toxicity of sunscreens and their degradation products on other reef members and partners that are key to overall reef health. Coral reefs are complex ecosystems that rely on a multitude of mutualistic interactions among diverse micro- and macrofauna inhabitants within reefs, including crustose coralline algae, sea anemones, and sponges (*12*). Thus, risk and toxicology assessments should incorporate feedback of the toxic effects between reef members and work to scale from the individual to the entire ecosystem.

In the face of a changing ocean, even perceived small actions may have compounding effects within compromised ecosystems. Because ecotourism is more popular than ever, efforts should be made to minimize the inadvertent contribution of sunscreen-derived toxic chemicals to the decline of reef health. The research presented by Vuckovic *et al*. provides critical insight for developing ecologically friendly sunscreens and should inform policy decisions for regulating sunscreen use within reefs and sensitive coastal ecosystems globally. ■

REFERENCES AND NOTES

1. O. Hoegh-Guldberg, L. Pendleton, A. Kaup, *Reg. Stud. Mar. Sci.* **30**, 100699 (2019).
2. O. Hoegh-Guldberg *et al.*, *Science* **318**, 1737 (2007).
3. M. Moeller *et al.*, *Front. Mar. Sci.* **8**, 665548 (2021).
4. D. Vuckovic *et al.*, *Science* **376**, 644 (2022).
5. A. Levine, *Mar. Policy* **117**, 103875 (2020).
6. R. Danovaro *et al.*, *Environ. Health Perspect.* **116**, 441 (2008).
7. C. A. Downs *et al.*, *Arch. Environ. Contam. Toxicol.* 70, 265 (2016).
8. M. J. H. van Oppen, L. L. Blackall, *Nat. Rev. Microbiol.* **17**, 557 (2019).
9. M. Stat, D. Carter, O. Hoegh-Guldberg, *Perspect. Plant Ecol. Evol. Syst.* **8**, 23 (2006).
10. T. P. Hughes *et al.*, *Science* **359**, 80 (2018).
11. C. L. Mitchelmore, E. E. Burns, A. Conway, A. Heyes, I. A. Davies, *Environ. Toxicol. Chem.* **40**, 967 (2021).
12. J. J. Stachowicz, *Bioscience* **51**, 235 (2001).

10.1126/science.abo4627

CORONAVIRUS

Rapid response modeling of SARS-CoV-2 transmission

What can modelers learn from recent history to help prepare for the next pandemic?

By **Jon Zelner**[1,2] *and* **Marisa Eisenberg**[1,3,4]

The COVID-19 pandemic has cemented the role of mechanistic infectious disease models as drivers of the scientific, public, and policy discourse during infectious disease emergencies. On page 596 of this issue, Pulliam *et al.* (*1*) add to these contributions through their use of a mechanistic model to document the high rate of reinfection with the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) Omicron variant in South Africa among people previously infected by the initial wild-type strain or the Alpha, Beta, or Delta variants. This work provides another example of how rapid-response modeling has facilitated the testing of key hypotheses and assumptions with unprecedented speed and near-immediate public health impact.

That sophisticated mechanistic models were rapidly pressed into action reflects decades of investment in the intellectual and technological resources to do so. It also reflects lessons learned from previous crises, including the 2003 SARS-CoV-1 outbreak, 2009 H1N1 influenza pandemic, and 2014 Ebola epidemic (*2*). Because modeling is inherently an integrative and historical enterprise, the failures of modeling during the COVID-19 crisis also reflect lessons not learned from these previous emergencies as well as ones that were impossible to anticipate owing to the novelty of the COVID-19 pandemic.

Much of the power of transmission models comes from their ability to create "stylized

[1]Department of Epidemiology, University of Michigan School of Public Health, Ann Arbor, MI, USA. [2]Center for Social Epidemiology and Population Health, University of Michigan School of Public Health, Ann Arbor MI, USA. [3]Department of Mathematics, University of Michigan, Ann Arbor, MI, USA. [4]Center for the Study of Complex Systems, University of Michigan, Ann Arbor, MI, USA. Email: jzelner@umich.edu

facts" about transmission and intervention. These are qualitative "rules of thumb" that flow from the quantitative output of mechanistic models (see the figure). It is these qualitative understandings that drive most of the scientific, political, and social conversation around infectious diseases in moments of deep uncertainty. When these accurately describe the state of the world, such stylized facts are invaluable. When they do not, there are serious consequences for public health and the public's confidence in it.

In the earliest moments of the pandemic, when data were sparse, modelers reached back to the 2003 SARS outbreak to understand whether it might be feasible to stem the global spread of this new—but related—pathogen (*3*). As data from Wuhan and other early outbreaks, such as on the Diamond Princess cruise ship (*4*), became available, analysts began to estimate the value of the basic reproduction rate R0 (the number of cases generated by one case in a susceptible population) and its responsiveness to intervention through the time-varying reproduction rate, R(t). One such analysis found that transmission rates in Wuhan declined over the course of early 2020, with a decisive dip that brought R(t) below the critical threshold of 1 after the institution of mobility restrictions in February 2020 (*5*). This and other analyses from the same period (*6*) suggested that lockdowns and other drastic contact-limiting interventions could be effective in stemming the tide of transmission. These insights shaped much of the global response to SARS-CoV-2 in 2020.

Whereas early analyses were characterized by a paucity of data, more information has become available with each wave of the pandemic to suggest what might happen next. For example, in a recent meta-analysis, the COVID-19 Forecasting Team used data from more than 3000 serological surveys conducted in 76 countries to characterize age- and region-specific infection fatality rates (IFRs) associated with SARS-CoV-2 infection (*7*). This type of "historical" data is critical for parameterizing models that can meaningfully capture spatial and demographic variation in the risks of infection and death. Such information is critically important for capturing the complex landscape of SARS-CoV-2 population immunity as the slow transition to endemicity continues (*8*).

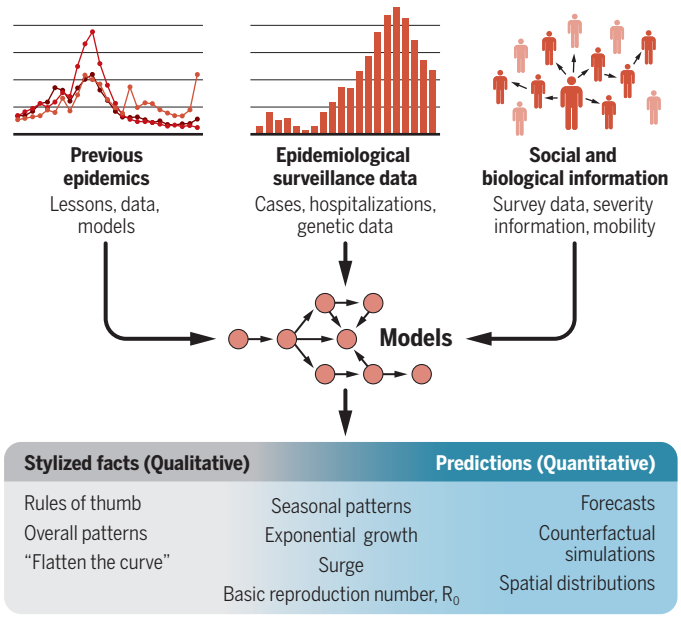
As the amount of SARS-CoV-2 data has increased, newly entrenched stylized facts that shape policy and behavior may make the response to shifts in the pandemic less nimble. These can ossify into "common sense" that provides a sense of control and predictability while actually crowding out important new information (*9*). For example, a consensus that emerged from years of research and discussion that respiratory viruses did not reliably transmit over distances greater than the 6-foot buffer used in social distancing guidelines was used to dismiss compelling observational evidence of an important role for longer-range airborne transmission of SARS-CoV-2 (*10*). As the volume of stylized facts about SARS-CoV-2 transmission rates, vaccine effectiveness, and IFRs continues to grow, the risk that new pieces of conventional wisdom will become disconnected from fast-changing realities increases.

Mental models also fail when they narrow the set of questions modelers think to ask: The role of racial discrimination and economic exploitation in amplifying the toll of COVID-19 among those who are financially and socially marginalized, and among people of color subjected to racial discrimination, was clear from the earliest moments of the pandemic (*11*). But the mechanistic models used as tools of rapid response at this time were focused largely on the pathogen (how infectious is it?) and not the brutally unequal human systems governing who is exposed and when (*12*). This has sparked a wave of new models that directly incorporate the mechanistic drivers of social inequality in infection and death (*13*). These allow quantification of the impact of social and economic policies on the burden of disease (*14*). In the future, the rapid deployment of high-quality models that adequately account for the joint social and biological drivers of infection will be critical to anticipating and addressing these inequities. This requires an immediate and concerted effort to increase socioeconomic, racial, and geographic diversity in the modeling workforce.

Mechanistic models have also played a contentious role in forecasting disease outcomes over the short to medium term. Although the predictive quality of disease forecasting has increased in recent years, these models necessarily reflect history-based assumptions about future transmission. As a result, even "ensemble" forecasting approaches that use multiple models with differing assumptions are often blindsided by rare, highly consequential events such as the emergence of SARS-CoV-2 variants with differing transmission characteristics, or rapid shifts in human behaviors, such as vaccination, masking, reduced air travel, and social distancing (*15*). Splitting the difference between using as much information as possible to suggest what might happen next, while acknowledging and quantifying the myriad deep uncertainties associated with a global pandemic of an evolving pathogen, must be a top priority of preparedness for future global infectious disease emergencies. ■

Modeling epidemics

Epidemiological models translate quantitative data and assumptions about the nature of infectious diseases into both qualitative "stylized facts" and quantitative predictions that have proven to be critical drivers of the public, political, and scientific conversation around severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2).



GRAPHIC: KELLIE HOLOSKI/*SCIENCE*

REFERENCES AND NOTES

1. J. R. C. Pulliam *et al.*, *Science* **376**, eabn4947 (2022).
2. J. Lessler, D. A. T. Cummings, *Am. J. Epidemiol.* **183**, 415 (2016).
3. C. Fraser, S. Riley, R. M. Anderson, N. M. Ferguson, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 6146 (2004).
4. J. Rocklöv, H. Sjödin, A. Wilder-Smith, *J. Travel Med.* **27**, taaa030 (2020).
5. A. J. Kucharski *et al.*, *Lancet Infect. Dis.* **20**, 553 (2020).
6. S. Flaxman *et al.*, *Nature* **584**, 257 (2020).
7. COVID-19 Forecasting Team, *Lancet* **399**, 1469 (2022).
8. J. S. Lavine, O. N. Bjornstad, R. Antia, *Science* **371**, 741 (2021).
9. T. Greenhalgh, *Interface Focus* **11**, 20210017 (2021).
10. L. Hamner *et al.*, *MMWR Morb. Mortal. Wkly. Rep.* **69**, 606 (2020).
11. W. N. Laster Pirtle, *Health Educ. Behav.* **47**, 504 (2020).
12. J. Zelner *et al.*, *PLOS Comput. Biol.* **18**, e1009795 (2022).
13. K. C. Ma *et al.*, *eLife* **10**, e66601 (2021).
14. A. Nande *et al.*, *Nat. Commun.* **12**, 2274 (2021).
15. E. Y. Cramer *et al.*, *medRxiv* 10.1101/2021.02.03.21250974 (2021).

ACKNOWLEDGMENTS

The authors thank K. Broen, R. Naraharisetti, and R. Trangucci for feedback on this manuscript.

10.1126/science.abp9498

MICROBIOLOGY

Bacterial species rarely work together

Competition is prevalent and could be harnessed as an alternative to antibiotics

By **Jacob D. Palmer**[1,2] *and* **Kevin R. Foster**[1,2]

Bacteria commonly live in diverse communities where each species can affect the growth and survival of others (*1*, *2*). These species interactions are central to bacterial ecology and have diverse implications for health, agriculture, and industry. To understand and manipulate bacterial communities, it is therefore critical to know how species interact. There is an emerging controversy about the importance of positive interactions, such as cooperation (or mutualism), between bacterial species (*3*–*8*). However, once the standard ecological measures are applied, a clear picture emerges. Negative interactions prevail, and cooperation, where two species both benefit, is typically rare. The prevalence of competition gives hope for bacterial community engineering strategies that seek to eliminate pathogens without the need for antibiotics.

Species affect each other in many ways. Antagonistic interactions such as competition, parasitism, and predation are rife with conflicts. An owl benefits from eating a mouse, whereas the mouse benefits if not eaten. Cooperative interactions, where both parties benefit, are characterized by natural selection on individuals of each species to work together. Plants disperse pollen through the flight of the bumble bee, whereas the bee obtains nectar through the plant's ability to fix atmospheric carbon. Species interactions also define how communities behave as a system, influencing diversity, assembly, stability, and productivity. For example, cooperation (+/+) can destabilize ecological systems by causing dependencies, whereas exploitation (+/−) can be stabilizing because it creates negative feedback between species (*9*). Species interactions are therefore critical to understanding complex communities.

Among interactions, cooperation can seem puzzling (*2*, *10*): Why benefit another over oneself? A vast literature is dedicated to this question, because answering it is central to explaining both complex life and societies (*10*). For bacteria and other microbes, cooperation readily evolves in clonal groups because of kin selection. When cells are genetically identical, natural selection can readily favor one cell investing in others, provided that the total reproduction of the group increases (*2*). However, between species, cooperation is predicted to be much less common because of natural selection for competition and exploitation when different genotypes interact (*1*, *2*, *10*).

In principle, this prediction can be tested using genomic data and computational methods that infer how different bacterial species interact (*2*). However, the gold standard is to culture species alone and together and directly measure their effects on each other. A decade ago, one study did such culturing with environmental bacteria isolated from tree holes (permanent rainwater pools in a beech tree woodland) (*8*). It was concluded that cooperation was uncommon, just as theory predicts. However, it was not clear whether this finding would be robust to different environments and methodologies (*3*). Recent work has provided experimental data on bacterial interactions from a range of environments, including the mammalian gut, nematode gut, phyllosphere (plant leaves and aerial tissue), and soil. How do these new data compare with the tree hole bacteria study? Most studies closely align with the tree hole bacteria data, showing that cooperation is rare (*4*–*7*). However, one concludes instead that positive interactions are common among bacterial species (*3*).

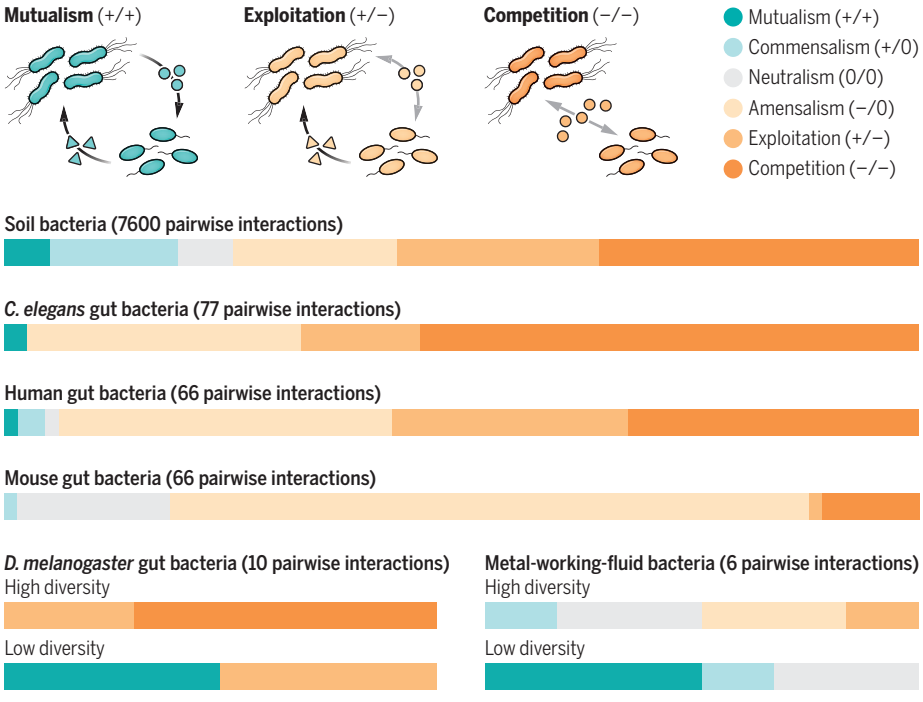
What is the reason for such different conclusions? One study tested a set of 12 bacterial species that colonize the mouse gut [Oligo-Mouse-Microbiota (OMM^{12})] (*4*). Pairwise interactions were quantified in vitro in nutrient-rich media. The data support the prediction that cooperation is rare: Most common was ammensalism (−/0; that is, negative in one direction), followed by neutral (0/0) and competitive (−/−) interactions. Indeed, of the 66 interactions measured, none were cooperative, and just one was commensal (+/0) (see the figure). Another in vitro study of human gut bacteria also found that negative interactions dominate (*7*). There were differences, with many fewer neutral (0/0) and more exploitative (+/−) interactions than in the OMM^{12} study. These shifts in estimates underscore the reality that different studies can yield quite different results, which may be caused by different methodologies as much as real differences between communities. Nevertheless, both of these in vitro studies support the idea that cooperation is rare, with it being absent in the OMM^{12} experiments and accounting for ~2% of interactions in the study of human gut bacteria. This rarity is also supported by in vivo experiments with bacteria associated with nematodes and plants. Pairwise experiments of bacteria from nematode guts revealed that competition was by far the dominant interaction, with cooperation again the rarest at ~3% of interactions (*5*). The plant study was carried out differently: Rather than comparing strains alone and in pairs, single species were removed or added in a diverse community. With this method, only ecological effects in one direction were assessed, but even so, negative interactions again predominated (*6*).

This is where things become less clear. Droplet-based culturing was used to test interactions between 20 species of soil bacteria (*3*). With this method, 190 species interactions across 40 different environments (7600 interactions in total) were surveyed, and >40% of the cases tested contained positive interactions. This result was contrasted against the environmental data from tree hole bacteria (*8*), where the authors of the new study noted that evidence of positive interactions was found in <10% of pairs of bacteria (*3*). With such a shift in the estimates, these data appear to suggest that soil communities are very different to those of tree holes and other environments. However, the soil bacteria study made a subtle but critical change in how "positive interactions" are defined, combining all cases with at least one positive interaction (*3*). This is problematic, because it groups exploitative interactions (+/−), such as parasitism and predation, with cooperation (+/+), which have fundamentally different evolutionary and ecological properties (*2*). Moreover, as a result of the change in definition, the suggestion that the tree hole bacteria study (*8*) had found evidence of positive interactions in <10% of cases becomes inaccurate because that estimate concerned solely cooperation. The importance of this distinction is made clear by looking at the frequency of cooperation proper (+/+) in the soil bacteria data, where it accounts for only 5% of interactions (*3*). This estimate is actually lower than that from the tree hole bacteria study.

[1]Department of Zoology, University of Oxford, Oxford, UK.
[2]Department of Biochemistry, University of Oxford, Oxford, UK.
Email: jacob.palmer@zoo.ox.ac.uk; kevin.foster@zoo.ox.ac.uk

Bacterial interactions in different communities

There are diverse interactions between bacterial species in a community, including mutualism (+/+), exploitation (+/−), and competition (−/−). How bacteria interact is important for community stability and composition. Analyses of the ecological interactions between bacterial species in soil (*3*), *Caenorhabditis elegans* gut (*5*), human gut (*7*), mouse gut (*4*), *Drosophila melanogaster* gut (*13*), and metal-working fluid (*12*) reveal that antagonistic interactions such as competition and exploitation dominate, whereas mutualism (cooperation) is rare. Examples of mutualism are seen in some studies when pairs of species are cultured ("low diversity") (bottom), but these disappear in cultures at more natural levels of diversity ("high diversity").



Whichever definition of positive ecological interactions is used, negative interactions are much more common than positive ones in the soil bacteria study. Such statistics do not mean, of course, that cooperation never evolves between bacterial species. There is evidence that cooperation has evolved between bacterial species of the human gut (*11*), even if it is generally uncommon. Moreover, one study found several cases where pairs of species grew better together than alone in toxic metal-working fluid (*12*). This study is important because it shows that the likelihood of bacterial species working together can greatly increase in certain environments. In this case, low species diversity and the challenges associated with growing in a harsh environment were found to be crucial. When nutrients or additional species were added, the cooperative interactions were lost (*12*) (see the figure). Consistent with the importance of diversity, a study of bacteria in the *Drosophila melanogaster* gut also found several positive interactions when species were cultured in pairs, but these nearly all shifted to negative when species diversity was increased to more natural levels (*13*). Indeed, the original tree hole bacteria study also found that the impacts of negative interactions increased with species diversity (*8*), presumably because the presence of more species increases the competition for limiting nutrients (*12*).

There is a need for further surveys of species and environments to understand the general characteristics of bacterial communities. Nevertheless, the emerging pattern is that cooperation is relatively rare, with negative interactions predominating. Another important pattern in the data is variability in the strength of ecological interactions, with many being relatively weak (*3*, *4*). Like negative interactions, weak interactions are predicted by evolutionary models: Strong competition between species is predicted to drive the extinction of one species or the evolution of niche separation (character displacement) that weakens the competition (*2*). However, these weakly negative interactions also present something of a conundrum. Many bacteria use powerful antibacterial toxins against one another, which create strong negative interactions (*1*). How are these observations reconciled? Antibacterial toxins often target members of the same species, which vie for the same nutrients and locations. Many strongly negative interactions, therefore, take place between different genotypes (strains) of one species, which are not captured by most ecological surveys that focus on different species. These intraspecific interactions deserve more attention and are needed for any complete picture of bacterial community ecology.

The emerging patterns in bacterial ecology have implications for those that seek to manipulate and engineer microbial communities for human benefit. Negative and weak interactions can be desirable from a community engineering standpoint because both can promote stability (*9*). Indeed, a key feature of the human gut microbiome is its relative stability, which allows it to recover from perturbations, such as a course of antibiotics (*9*). Negative interactions can also drive priority effects, where late-arriving species are unable to grow because of the effects of early arriving species (*6*). Competition and exploitation, therefore, can make the addition of new species challenging. This effect is well known from studies of probiotic use, where "beneficial" bacterial species often fail to colonize the gut (*14*). However, such colonization resistance can also be a benefit when incoming bacteria are harmful. With rising levels of drug resistance in pathogenic bacteria, alternatives to antibiotics are urgently needed. After the success of fecal transplants for the treatment of *Clostridium difficile*, there is great interest in finding bacteria that both colonize the human microbiome and compete strongly with pathogens (*15*). If bacterial community ecology can be mastered, the hope is that these competitive species can be introduced as a prophylactic, or even as a treatment, to eliminate pathogens. The way out of the crisis in antibiotic resistance may rest upon both the prevalence and power of bacterial competition. ■

REFERENCES AND NOTES

1. M. Ghoul, S. Mitri, *Trends Microbiol.* **24**, 833 (2016).
2. S. Mitri, K. R. Foster, *Annu. Rev. Genet.* **47**, 247 (2013).
3. J. Kehe *et al.*, *Sci. Adv.* **7**, abi7159 (2021).
4. A. S. Weiss *et al.*, *ISME* **6**, 1095 (2021).
5. A. Ortiz *et al.*, *ISME J.* **15**, 2131 (2021).
6. C. I. Carlström *et al.*, *Nat. Ecol. Evol.* **3**, 1445 (2019).
7. O. S. Venturelli *et al.*, *Mol. Syst. Biol.* **14**, e8157 (2018).
8. K. R. Foster, T. Bell, *Curr. Biol.* **22**, 1845 (2012).
9. K. Z. Coyte *et al.*, *Science* **350**, 663 (2015).
10. S. A. West *et al.*, *Curr. Biol.* **17**, R661 (2007).
11. S. Rakoff-Nahoum *et al.*, *Nature* **533**, 255 (2016).
12. P. Piccardi *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 15979 (2019).
13. A. L. Gould *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **115**, E11951 (2018).
14. M. Derrien, J. E. T. van Hylckama Vlieg, *Trends Microbiol.* **23**, 354 (2015).
15. M. T. Sorbara, E. G. Pamer, *Mucosal Immunol.* **12**, 1 (2019).

ACKNOWLEDGMENTS

We thank A. Ågren, B. Klosterhoff, S. Mitri, and members of the Foster lab for helpful comments. K.R.F. is funded by European Research Council grant 787932 and Wellcome Trust Investigator Award 209397/Z/17/Z.

10.1126/science.abn5093

GRAPHIC: A. MASTIN/*SCIENCE*

POLICY FORUM

RIVER GOVERNANCE

Save the Mekong Delta from drowning

Policy must address drivers, not just symptoms, of subsidence

By **G. M. Kondolf[1], R. J. P. Schmitt[2,1], P. A. Carling[3], M. Goichot[4], M. Keskinen[5], M. E. Arias[6], S. Bizzi[7], A. Castelletti[8], T. A. Cochrane[9], S. E. Darby[3], M. Kummu[5], P. S. J. Minderhoud[10,11,12], D. Nguyen[13], H. T. Nguyen[14], N. T. Nguyen[15], C. Oeurng[16,1], J. Opperman[17], Z. Rubin[1,18], D. C. San[19], S. Schmeier[20], T. Wild[21]**

Owing to only a few decades of human influence and unsustainable management of the Mekong River basin's natural resources, the Mekong Delta is receding rapidly. Most of the delta landform, home to 17 million people and an economic powerhouse, could slip below sea level by 2100 (*1*). Avoiding such a catastrophic impact will require concerted actions that acknowledge root causes for land loss and the global importance of the delta landform. Deltas persist and grow if sediment supply from an upstream river basin builds delta land at the same or greater rates than land is submerged by relative sea level rise and erosion. With more rapid sea level rise, more sediment resources are needed to maintain the current extent of the delta. Only improved coordination of governance and investments, informed by science, will provide the delta with those critical resources.

The Mekong Delta, which lies mostly in Vietnam, is one of the world's largest deltas. The delta has been transformed in the past century into a human-made landscape, or "Delta Machine" (*2*), which produces 7 to 10% of all rice traded internationally. The delta averages less than 1 m above sea level, so it is vulnerable to subsidence and coastal erosion. Many initiatives have supported local adaptation measures to address symptoms of a sinking delta, but have not addressed underlying anthropogenic drivers of subsidence at both the delta and basin scales, nor considered the international nature of the basin.

Prior to the late 20th century, the delta received 140 to 160 million metric tons (Mt) of sediment annually from the Mekong River basin. More than half of this is now being trapped in reservoirs. In the upper Mekong basin in China (the Lancang), eight large hydropower dams have been completed, with another 20 under construction or planned. In the Mekong, 133 dams are built or planned, of which 11 are on the mainstem of the lower Mekong. If built as planned, all dams will trap 96% of the sediment formerly reaching the delta (*3*). Additionally, sediment supply from tropical cyclones, which deliver about 32% of the suspended sediment load reaching the delta, is decreasing as cyclone tracks shift north (*4*).

The remaining sediment load is further reduced by in-channel mining. An estimated 54 Mt of sand per year from the Mekong River, mostly in Cambodia and Vietnam, is used in construction and land reclamation (*5*). Sand mining causes downstream sediment starvation and contributes to coastal erosion and channel incision, tidal amplification, and salinity intrusion (*6*).

Management of the delta has historically focused on controlling the waters to enable agricultural intensification and flood regulation, and to prevent saline water intrusion. Although successful in this regard, this has fundamentally affected natural processes that maintain the delta land itself. Where distributary channels and coastal currents formerly distributed sediment-laden flood flows across the delta plain and along its coastlines, dikes now restrict water and sediment to the main channels, depriving the delta of deposition during floods. Natural mangrove vegetation traps sediment to build up land, absorbs wave energy, and reduces coastal erosion. However, the delta's mangroves largely have been replaced by agriculture and aquaculture, and remaining mangroves are now starved of sediment to trap (*7*).

All deltas naturally subside, as recently deposited sediment compacts. For the Mekong Delta, this natural subsidence is exacerbated by effects of groundwater pumping for agriculture and urban use, presently the single greatest driver of subsidence in the delta (*8*). By 2100, a "business as usual" scenario results in average relative subsidence of up to 1.8 m, which would lead to submergence of over 90% of the delta. A best-case scenario (strongly curtailed pumping, mining, and dam construction) results in subsidence of 0.15 m, which would inundate about 10% of the delta (*1*).

The above-mentioned drivers can create vicious cycles. For example, as salt water intrudes into the delta, farmers may use more groundwater, or migrate to urban centers that are already foci of subsidence. As subsidence accelerates, building dikes to lock out floods becomes more attractive to local interests, but these dikes prevent sediment from spreading over the delta surface and building elevation.

ADDRESSING ROOT CAUSES OF SUBSIDENCE

The very existence of the Mekong Delta as we know it today is due to massive human-made modifications—canals, dikes, saltwater dams, and other hydraulic interventions—that have led to major ecological and economic transformations (*9, 10*). Since the Vietnamese reunification in 1975, the delta has seen a number of studies and master plans, most with international support, to promote centralized, integrated planning with focus on socioeconomic development (*11*). Though highly successful in turning the delta into an agricultural and economic powerhouse, this has increasingly locked the delta's management into an unsustainable path with weak adaptation capacity, siloed governance, and lack of coordination with actions in upstream countries.

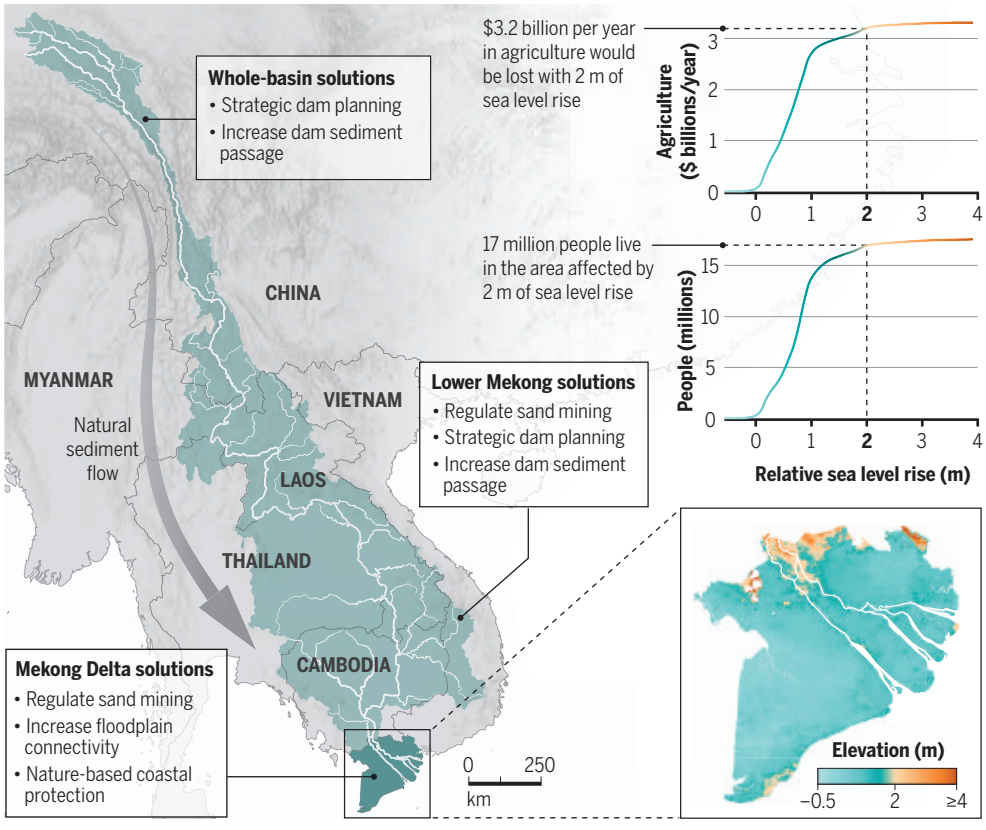
[1]Riverlab, Department of Landscape Architecture and Environmental Planning, University of California, Berkeley, CA, USA. [2]The Natural Capital Project, Stanford University, Stanford, CA, USA. [3]School of Geography and Environmental Science, University of Southampton, Southampton, UK. [4]World Wide Fund for Nature Asia Pacific, Ho Chi Minh City, Vietnam. [5]Water and Development Research Group, Aalto University, Espoo, Finland. [6]Department of Civil and Environmental Engineering, University of South Florida, Tampa, FL, USA. [7]Department of Geosciences, University of Padova, Padua, Italy. [8]Department of Electronics, Information, and Bioengineering, Politecnico di Milano, Piazza Leonardo da Vinci, Milano, Italy. [9]Department of Civil and Natural Resources Engineering, University of Canterbury, Christchurch, New Zealand. [10]Soil Geography and Landscape group, Wageningen University, Netherlands. [11]Department of Civil, Environmental and Architectural Engineering, University of Padova, Padova, Italy. [12]Subsurface and Groundwater Systems Unit, Deltares Research Institute, Utrecht, Netherlands. [13]Laboratory for Hydraulics Saint-Venant, Université PARIS-EST, Chatou, France. [14]Can Tho, Vietnam. [15]University of Science, Vietnam National University, Ho Chi Minh City, Vietnam. [16]Institute of Technology of Cambodia, Phnom Penh, Cambodia. [17]Global Science, World Wildlife Fund, Washington, DC, USA. [18]Balance Hydrologics, Berkeley, CA, USA. [19]Southern Institute of Water Resources Research, Ho Chi Minh City, Vietnam. [20]Water Governance Department, IHE Delft Institute for Water Education, Delft, Netherlands. [21]University of Maryland, College Park, MD, USA. Email: kondolf@berkeley.edu

The focus on socioeconomic development and strong belief in human mastery over nature also explains why current policies fail to properly acknowledge that most of the delta might fall below mean sea level within a human lifetime. The key Vietnam government document guiding the delta's development, Resolution 120 (enacted in 2017), takes steps in the right direction, emphasizing the need for more nature-based development and encouraging integrated planning across different sectors and spatial scales, and aiming to "develop an integrated plan for sustainable and climate resilient development of the Mekong Delta" (*12*). Yet, the Resolution and the earlier "Mekong Delta Plan" (*13*) treat flooding, salinity, and coastal erosion more as isolated engineering challenges, proposing solutions on local scales, rather than explicitly seeing them as symptoms of underlying causes spanning multiple scales. Although the Plan reports ongoing subsidence rates, the existential threat to the delta is not addressed by the limited proposals to minimize subsidence, reestablish sediment connectivity, and increase delta resilience. This gap in management is also not overcome in more recent plans, such as "The Mekong Delta Integrated Regional Plan" proposed to the Vietnamese Ministry of Planning and Investment in July 2020 or the government report reviewing the first 3 years of the implementation of Resolution 120 (*14*).

At the same time, large-scale investments keep on pouring into the delta, with a mixture of instruments to advance development and respond to the threat posed by climate change. For example, the World Bank has committed to loans and bilateral funding of nearly US$2 billion from 2007 to 2022. Although this level of funding is small compared to private investment, funding decisions of international financial institutions have tremendous influence in setting standards for other investors and could help push investments that make the delta more resilient. Reputational risks are an increasingly important motivator for private investors, who are influenced by the examples set by major international actors. Much of the funding from international donors aims to address the impacts that a sinking delta creates on the livelihoods of its inhabitants: to protect against local flooding and coastal erosion, deal with salinized water supplies, and support adaptation planning. However, the investment proposals typically say little or nothing about the long-term existential risks to both livelihoods and investments, nor about opportunities to mitigate those trends through concerted larger-scale actions.

The drowning Mekong Delta

The existence of the Mekong Delta is threatened by anthropic drivers. Continuing unsustainable use of water and sediment will result in the delta surface dropping by up to 2 m relative to sea level by 2100, drowning most livelihoods and ecosystems in the delta. Measures to avoid this fate are known but need to be implemented urgently across scales, sectors, and borders (see supplementary methods for data sources).



GRAPHIC: K. FRANKLIN/*SCIENCE*

RESPONDING TO THE THREATS

Persistent challenges from top-down, hierarchical governance systems and institutional rivalries hinder continued positive evolution. For the delta landform to persist into the future, three fundamental changes in the political economy and science-society-policy nexus are required to enable six deep changes in basin management, with particular attention to sediment.

First, investments in the delta across key economic sectors (agriculture, aquaculture, transport, energy, construction) and actors (national and international companies, state-owned enterprises, local and national governments, and international development actors) need to register and address system-scale consequences of their investments, such as their impact on delta subsidence. This would require improved planning mechanisms, based on enforced regulatory frameworks. Second, actors such as civil society groups and regional and international investors with an interest in sustainable and long-term profitable investments must leverage their shared interest in making basin-scale change. For example, they can highlight that avoiding high-impact dams will benefit not only local communities but also the delta and entire basin. Third, science–policy interaction must be enhanced. One problem in Mekong basin science has been that minor differences among scientific studies have been emphasized. This gives the impression of great uncertainty in scientific predictions, but in reality, there is broad scientific consensus on key concerns. These should be emphasized when scientists communicate with decision-makers. Decision-makers, in exchange, should not take remaining differences in scientific findings as an excuse for inaction.

To respond to the fundamental threat facing the delta, future projects should not contribute to increased subsidence and land loss, and they should be resilient to future subsidence. To achieve this, we recommend that future measures explicitly account for the scales at which the drivers operate, and that management efforts be undertaken at appropriate scales. Although some drivers of subsidence can be influenced by local and national actions, others will require interventions coordinated between Cambodia and

Vietnam, and throughout the entire basin, from which the essential water and sediment resources are derived.

Coordinated planning on larger scales might seem unlikely given current political realities and persistent governance challenges. Changing those realities toward a whole-basin dialogue is needed to open a larger solution space where all riparian countries acknowledge the situation and have both rights and obligations. Here, active science–policy interaction, including the role of knowledge intermediaries such as regional networks and civil society organizations (*2*, *10*, *15*), plays an important role. Scientific understanding of the existential threats facing the delta is clear, and this understanding must be actively translated into policy recommendations. Translating such recommendations into impactful decisions requires political will and coordinated actions, both at national and basin scales. This requires accepting that some critical activities, such as overpumping, sand mining, and hydropower development, may need to be limited or gradually phased out, despite their political importance and economic interests.

In this context, we propose six measures to safeguard the delta and its livelihoods. We identify potential enablers and roadblocks to implementation (table S1). The measures will be most effective in concert, and implementing the full portfolio will be a challenge. Yet each measure has precedents (table S1).

Avoid high-impact dams

Do not build dams at sites with the highest sediment-trapping potential. Use network-scale portfolio analysis to identify optimal dam placement to minimize impacts while maintaining hydropower production. The benefits and impacts of alternative energy sources should be considered along with hydropower.

Pass sediment through or around dams

Sustainable sediment management strategies such as sluicing, flushing, and bypasses can allow some sediment to move from upstream to downstream, as included in the mainstem dam design guidance of the Mekong River Commission (MRC). Studies are needed to evaluate if dams currently not equipped for sediment management can be retrofitted.

Phase out riverbed sand mining

Impacts of sand mining can be reduced through better enforcement of mining regulations, limiting extraction to a sustainable fraction of remaining sediment load, and encouraging alternative material for construction, e.g., from floodplains and recycled material.

Transform agriculture in the delta

Produce less quantity but higher quality, and adapt agricultural practices to minimize groundwater extraction and reconnect distributaries to the delta plain. As recommended by the Vietnam Government Resolution 120 (*12*), reduce groundwater pumping through a shift to less–water-intensive crops, improve access to surface water supplies and maintain their quality, regulate water demand, and reuse water.

Maintain connectivity of delta floodplains

Adapt water infrastructure in the delta to allow sediment-charged flows to spread over the delta surface at least one year in three, and acknowledge the socioeconomic benefits of floods and sediment. Prevent dikes from cutting off channels from floodplains.

Leverage nature-based coastal protection

At low cost, mangroves and natural wetlands are a proven solution to coastal erosion with benefits for biodiversity and livelihoods, but for those ecosystems to build land there must be a supply of sediment for them to trap.

SEVERE, URGENT, EXISTENTIAL

By virtue of its national, regional, and global importance, the Mekong Delta has attracted extensive interest from the Vietnamese government, international development partners, multinational corporations, researchers, and civil society (*2*, *10*, *15*). Yet the severity and urgency of the existential threat—that most of the delta will sink below sea level by 2100—have not been explicitly mainstreamed in key policies and investment plans. The six proposed measures will entail major costs and will require coordination between civil, business, and political actors to navigate trade-offs between sectors at national and regional scales.

The proposed six-measure portfolio will not be easy to implement, but above we outline some enablers that will be crucial in the basin's governance, private sector, academia, and civil society. In the delta, national and provincial governments could act now to develop and enforce regulations for sustainable groundwater management and invest in their implementation.

In the basin, existing transboundary organizations must play a key role. Although the MRC is unlikely to gain binding power to make strategic development decisions, it needs to strengthen its role as knowledge broker and amplifier of transboundary environmental issues (*15*), including demonstrating that cooperation yields greater benefits than unilateral action. As more investments will pour into the basin, e.g., with the intensification of ASEAN (Association of Southeast Asian Nations) power trade, such information will be crucial to hold private and government investors accountable for the externalities of their investments.

Organizations with active transboundary and trans-sectoral investment portfolios, e.g., multilateral development banks such as the World Bank and Asian Development Bank, are already positioned to evaluate the systemic impacts of investments rather than focusing only on project-level sustainability assessments (*10*). Because of the high visibility of multilateral development banks and other international organizations, their commitment to strategic and systemic solutions can further awareness among national actors, and they are well positioned to support sectoral reforms and highlight the role of strategic assessments and sustainable use of riverine resources.

All those organizations and actors need to broaden their scope from a narrow focus on developing the basin's water and energy resources (*10*) to consider additional sectors (e.g., construction, food) and holistic solutions to address root causes of a subsiding delta. Implementing those solutions could make the Mekong Delta a sentinel of integrative and innovative delta management and climate adaptation, rather than being drowned by the century's end. ■

REFERENCES AND NOTES

1. R. J. P. Schmitt *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2026127118 (2021).
2. F. Molle, T. Foran, M. Käkönen, *Contested Waterscapes in the Mekong Region: Hydropower, Livelihoods and Governance* (Earthscan, 2009).
3. G. M. Kondolf, Z. K. Rubin, J. T. Minear, *Water Resour. Res.* **50**, 5158 (2014).
4. S. E. Darby *et al.*, *Nature* **539**, 276 (2016).
5. J.-P. Bravard, M. Goichot, S. Gaillot, *EchoGéo* **26** (2013).
6. S. Eslami *et al.*, *Sci. Rep.* **9**, 18746 (2019).
7. M. Besset *et al.*, *Estuar. Coast. Shelf Sci.* **226**, 106263 (2019).
8. P. S. J. Minderhoud, H. Middelkoop, G. Erkens, E. Stouthamer, *Environ. Res. Commun.* **2**, 011005 (2020).
9. T. A. Tran, *Asia Pac. Viewp.* **61**, 162 (2020).
10. M. Käkönen, thesis, University of Helsinki, Helsinki, Finland (2020).
11. M. F. van Staveren, J. P. M. van Tatenhove, J. F. Warner, *J. Environ. Policy Plann.* **20**, 267 (2018).
12. Government of Viet Nam, Government Resolution 120/NQ-CP on Sustainable and Climate-Resilient Development of the Mekong Delta og Viet Nam (2017).
13. MoNRE, "Mekong Delta Plan" (Ministry of Natural Resources and Environment, Hanoi, Vietnam, 2013), p. 126.
14. VGP, "Preliminary Outcomes of Three-year Implementation of Resolution on Climate-Resilient Development of Mekong Delta" (2021).
15. N. T. Nguyen, in *Collaboration in Water Resource Management in Vietnam and South-East Asia* (Nomos, 2020), pp. 35–44.

ACKNOWLEDGMENTS

G.M.K. and R.J.P.S. contributed equally to this work.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm5176

10.1126/science.abm5176



BOOKS *et al.*

Once a lawless free-for-all, the US fishing industry is now tightly regulated.

MARINE ECOLOGY

The changing story of seafood

New technology, new markets, and better policy are improving fisheries and aquaculture

By **Olaf P. Jensen**

Farms and those who work them evoke amber waves of grain and *American Gothic* for many Americans, images compelling enough to make us forget the tallgrass prairie long ago plowed under to create them. Fishing and aquaculture are a different story, but which one—the ocean bottoms dredged barren by factory trawlers, or the cornucopia of seafood spilling from a Dutch still life? Without seeing seascapes for ourselves, we rely on storytellers to provide a narrative. The dominant ones today are defined by decline and destruction: disrupted food chains, farmed salmon pumped full of antibiotics, all fish gone by 2048.

In *The Blue Revolution*, Nicholas Sullivan provides a different, more nuanced narrative of US fishing. In the book's first section, he recounts the transition of wild-capture fisheries in New England from a lawless coastal version of the Wild West to the highly regulated system in place today. Until the mid-1990s, fishing operations competed in a "race to fish" free-for-all: the more you caught, the more you earned, and if you did not catch a fish, someone else would. Fish populations suffered the consequences.

Today, all aspects of a US fishing trip are tightly circumscribed. When, where, and how to fish are governed by regulation. Fishing operations must call in to let fishery managers know that they are starting a trip and then file reports on what they caught. The cowboys have been reined in by paperwork, and the outlaws—Sullivan focuses on the colorful "Codfather" Carlos Rafael, who bought up much of the New Bedford fishing fleet with proceeds from illegal cod—have been brought to justice.



The Blue Revolution
Nicholas P. Sullivan
Island Press, 2022.
272 pp.

A whole ecosystem of businesses connecting fish catchers to fish eaters has recently sprung up. Sullivan tells the stories of numerous entrepreneurs who have found a niche, marketing seafood to discerning buyers—both restaurants and retailers—who want their fish to be not only fresh but also sustainable and local.

Some fishing operations have cut out the middleman and are selling directly to consumers at roadside stands, at farmers' markets, or through pay-in-advance community-supported fisheries. Direct sales of seafood helped fishermen in New England weather the COVID pandemic (*1*), as the high-volume supply chain dried up and many consumers began experimenting with cooking locally caught species at home.

But how many people are willing to pay double for seafood of known provenance? Clearly enough, Sullivan argues, to support the rapid expansion of businesses such as Sitka Salmon Shares, which brings sustainable wild Alaskan seafood to dinner plates throughout the Lower 48. However, we have known for decades that the potential sustainable wild harvest from the world's oceans is ~100 million tons per year (*2*), and that is pretty close to where it seems to have leveled off since the late 1980s (*3*). The rise in global seafood production since then is attributable entirely to the growth in aquaculture, a trend that must continue if we are going to feed fish to the masses.

Aquaculture production in the US, and the technology that could help it expand, are the subject of the second section of Sullivan's book. This section is well sourced but gets bogged down in overwhelming detail. It is difficult to keep track of which fish farmer was childhood friends with another and who worked on someone else's aquaculture project before taking over as CEO of a kelp business. Here too, Sullivan's discussion of kelp farming and "integrated multitrophic aquaculture" is weakened by its wide-eyed belief in techno-utopian ideas that always seem to be poised to imminently save the planet.

It is not until the end of this section that Sullivan gets to the root of what has kept US aquaculture small: the lack of an efficient permitting process. Here, he quotes Scott Flood, an ocean engineer and lawyer, on the limitations of bluewater aquaculture: "It's all in the permitting, as the technology is pretty well understood…"

The role of policy, both in driving the revolution in wild-capture fishery sustainability and in hindering the growth in aquaculture in the US, is given short shrift in Sullivan's book. The Magnuson-Stevens Act amendments of 1996 and 2006, for example, required catch limits on all fisheries, made overfishing illegal, and required overfished stocks to be rebuilt by a certain deadline. These and similar measures implemented in other countries can be directly linked to reductions in fishing pressure and the recovery of fish populations (*4*).

To some extent, Sullivan's neglect of policy successes is understandable. A bill is an unlikely hero (*Schoolhouse Rock!* excepted), and a book about the triumph of well-crafted legislation is unlikely to appear on the best-seller list. However, in a book about fisheries that gets so much right, it is disappointing not to see credit given where it is due. ■

The reviewer is at the Center for Limnology, University of Wisconsin–Madison, Madison, WI 53706, USA. Email: ojensen@wisc.edu

REFERENCES AND NOTES

1. S. L. Smith, A. S. Golden, V. Ramenzoni, D. R. Zemeckis, O. P. Jensen, *PLOS ONE* **15**, e0243886 (2020).
2. J. H. Ryther, *Science* **166**, 72 (1969).
3. *The State of World Fisheries and Aquaculture 2020. Sustainability in Action* (Food and Agriculture Organization of the United Nations, 2020).
4. M. C. Melnychuk *et al.*, *Nat. Sustain.* **4**, 440 (2021).

10.1126/science.abo6915

PHOTO: MONTY RAKUSEN/GETTY IMAGES

ANTHROPOLOGY

Hidden influences

Beliefs about the vital roles played by spirits and gods have long been underappreciated, argues an anthropologist

The New Science of the Enchanted Universe
Marshall Sahlins, with the assistance of Frederick B. Henry Jr.
Princeton University Press, 2022. 208 pp.



By **Jonathan Spencer**

In the last decade of his long and extraordinarily creative life as an anthropologist, Marshall Sahlins, who died aged 90 in 2021, found a new voice on social media. On Facebook, he railed with his customary sharp wit, mostly against the state of US politics but also sometimes against the state of US anthropology. One of his posts, from 2017, which he titled "Where have all the cultures gone?" and described as an "emeritus rant," spread virally, generating endorsements and attacks in equal measure. In the post, Sahlins asked what had happened to anthropology as "the encompassing human science" and, more specifically, why had a new generation apparently turned its back on the accumulated knowledge of human diversity that had once been the discipline's stock in trade? *The New Science of the Enchanted Universe*, completed with the help of friends and family in Sahlins's final months, is in many ways the continuation of that argument, albeit recast in a more conventional scholarly form, with considerable erudition (and no little humor).



Trobriand Islanders view their ancestors' interventions as key to successful harvests.

The New Science, which takes its title in a mildly ironic homage to the philosopher Giambattista Vico, was originally conceived of as the introductory volume in a trilogy. The other two volumes, on "Enchanted Economics" and "Cosmic Politics," will presumably never appear. Although this is a loss, there is no doubt that Sahlins's *New Science* can stand alone as a characteristically feisty final statement from one of the greatest anthropologists of the past century.

The book is animated by a strong central premise: that most human beings for most of our known history inhabited cultural worlds in which nonhuman entities—spirits, gods, or Sahlins's preferred terms, "metapersons" and "metahumans"—played a constant and decisive part in human affairs. Our ability to acknowledge and understand this state of affairs is inhibited, often fatally, he argues, by the intellectual biases that have accompanied the world-historical shift from "immanent" understandings of divinity in the world to "transcendental" perspectives that displace divinity to another world with less direct engagement in everyday human affairs. That shift, which saw the birth of universalist religions, was characterized by the philosopher Karl Jaspers as the "Axial Age," a term that has spawned its own subfield in the history and sociology of religion.

Sahlins is not, however, especially interested in the historical and sociological arguments about the tension between immanent and transcendental tendencies in the history of the world religions. His use of the Axial Age argument is rather more limited and, one suspects, strategic: He uses it to group together examples from ancient history and more recent ethnography, from pre-Axial civilizations, and from classic accounts of what are now thought of as Indigenous peoples.

The ancestor whose spirit animates the bulk of the book is not Jaspers but rather the inventive but little-known British anthropologist-archaeologist A. M. Hocart. Indeed, the main lineaments of Sahlins's argument in *The New Science*, and many of the strongest examples he discusses, were set out earlier in a 2016 lecture, published as "The Original Political Society" (*1*). That lecture was the first—and thus far only—Hocart Memorial Lecture and was republished in a brilliant collection of essays on kingship, co-written with Sahlins's protégé, the late David Graeber (*2*).

Hocart had argued that humans lived with gods before they lived with kings and that the beginnings of complex administrative orders lay in complex ritual orders, which only later became adapted to the process we would recognize as government. This provides the blueprint for Sahlins's delineation of the main features of his enchanted universe, the world as lived and understood by most humans most of the time. Sahlins's account runs across four thematic chapters on human finitude, immanence, metapersons, and what he calls the "cosmic polity."

The book's examples come from classic work by earlier anthropologists and historical cases taken from Sahlins's wide and sometimes surprising reading, supplemented by more recent work from Melanesia and Amazonia. Those two regions have been the favored sites of what has become known as the "ontological turn" in recent anthropology, a call to recognize the strong ontological claims implicit in other peoples' accounts of their worlds. Sahlins echoes the common rhetorical claim in this more recent literature that other anthropologists fail to take other people's claims about the world seriously. He does not, however, follow some other anthropologists who have been gleefully finding evidence of immanent reenchantment in the most unlikely of modern settings (*3*).

Famous for his aphorisms, Sahlins once claimed there were two certainties: "'In the long run, we all die and we are all wrong. A good career is when the former happens before the latter." Readers will relish the opportunity that *The New Science* provides to test that generalization. ■

The reviewer is at the School of Social and Political Science, University of Edinburgh, Edinburgh EH8 9LD, UK. Email: jonathan.spencer@ed.ac.uk

PHOTO: GEORG BERG/ALAMY STOCK PHOTO

REFERENCES AND NOTES

1. M. Sahlins, *HAU* **7**, 91 (2017).
2. D. Graeber, M. Sahlins, *On Kings* (HAU Books, 2017).
3. B. Meyer, P. Pels, Eds., *Magic and Modernity* (Stanford Univ. Press, 2003).

10.1126/science.abo6935



LETTERS

Dense tree plantations are vulnerable to wildfires, as shown by this pine and spruce forest in Sweden, which burned in a 2014 megafire.

PHOTO: ALEXANDRO B. LEVERKUS

Edited by **Jennifer Sills**

Tree planting goals must account for wildfires

Grassroots movements such as the Trillion Tree Campaign (*1*) and international policies such as the EU Biodiversity Strategy for 2030 (*2*) aim to mitigate climate change through ambitious tree planting objectives. However, tree planting targets could produce counterproductive side effects (*3*), including an increase in the amount and continuity of fuels, a key driver of large fires in a warming world (*4*, *5*). Drier weather, coupled with afforestation dominated by extensive, dense, even-aged, monospecific conifer or eucalypt plantations, has already promoted megafires in places like Chile and Portugal (*6*), and burnt areas are more likely to reburn when postfire management includes extensive reforestation instead of natural regrowth (*7*). Reforestation programs should prioritize the mitigation of fire risks.

Strategies to combat climate change through tree planting—whether through plantations or through restoring native vegetation—should address how the changes in composition and configuration of landscapes can affect fire propagation. Plans should favor landscape mosaics, heterogeneous and fragmented stands rather than large and homogeneous ones, vertical vegetation discontinuities that prevent surface fires from spreading to crowns, high species diversity, low-flammability species, and low plant densities throughout the life span of planted trees (*8*). In addition, decisions about which tree species to plant should prioritize natural resilience to future fires, which are likely to occur at large spatial and temporal scales under warming conditions (*9*). In many places, native resprouting species would meet these criteria.

Many ongoing land management efforts aim to help adapt ecosystems to climate change by selecting species adapted to projected future climate (*10*). Preparing for the likely increases in frequency, severity, and extent of future wildfires is at least as important. Rather than targeting a specific number of trees, reforestation programs should account for factors like the potential for the planted trees to capture carbon in the long term, considering the influence of feedbacks between planting and altered fire patterns (*5*). Revegetation that does not consider these processes may accentuate the risk of megafires and thereby abruptly release large amounts of carbon, reduce vegetation cover, and increase the risk of ecosystem collapse.

Alexandro B. Leverkus[1]*, Simon Thorn[2], David B. Lindenmayer[3], Juli G. Pausas[4]
[1]Departamento de Ecología, Universidad de Granada, 18071 Granada, Spain. [2]Field Station Fabrikschleichach, Department of Animal Ecology and Tropical Biology, Biocenter, University of Würzburg, 96181 Rauhenebrach, Germany. [3]Fenner School of Environment and Society, The Australian National University, Canberra, ACT 2601, Australia. [4]Centro de Investigaciones sobre Desertificación–Consejo Superior de Investigaciones Científicas, 46113 Montcada, Valencia, Spain.
*Corresponding author. Email: leverkus@ugr.es

REFERENCES AND NOTES

1. Trillion Tree Campaign; www.trilliontreecampaign.org/.
2. European Commission, "Biodiversity strategy for 2030" (2020).
3. K. D. Holl, P. H. S. Brancalion, *Science* **368**, 580 (2020).
4. J. G. Pausas, J. E. Keeley, *Front. Ecol. Environ.* **19**, 387 (2021).
5. A. Duane, M. Castellnou, L. Brotons, *Clim. Change* 165, **43** (2021).
6. S. Gómez-González, F. Ojeda, P. M. Fernandes, *Environ. Sci. Pol.* **81**, 104 (2018).
7. J. R. Thompson, T. A. Spies, L. M. Ganio, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 10743 (2007).
8. A. B. Leverkus *et al.*, *Environ. Res. Lett.* **16**, 021003 (2021).
9. UN Environment Programme (UNEP), "Spreading like wildfire: The rising threat of extraordinary landscape fires" (Nairobi, 2022).
10. O. Hoegh-Guldberg *et al.*, *Science* **321**, 345 (2008).

10.1126/science.abp8259

Adapt biodiversity targets to climate change

Although the climate crisis is interrelated with biodiversity loss, the decade-old targets of the Convention on Biological Diversity (CBD) have barely addressed climate change impacts (*1*). So far, the post-2020 global biodiversity framework continues to miss opportunities: The first draft, released in July 2021 (*2*), overlooks climate-biodiversity interactions and provides no explicit solutions to anticipate climate change–related risks. These issues persist after the latest input of scientific experts (*3*, *4*). A post-2020 global biodiversity framework needs to include adaptation, not just mitigation, to achieve biodiversity goals by 2050.

Climate impacts (such as habitat fragmentation and ecological disruption) will escalate and interact with other destruction drivers (such as land degradation and overexploitation) to constrain ecosystems' integrity and functioning, which will threaten species survival globally (*5*). Disregarding these scenarios [as in table 1 and figure 1.1 in (*3*)] will likely compromise the CBD's efforts to pursue ambitious targets (e.g., protecting 30% of Earth's surface by 2030) and to expand conservation dimensions by finally safeguarding genetic diversity. To make targets climate-resilient, forward-looking strategies need to be developed. Primarily, the expansion of protected areas (target 3) should prioritize sites that can act as climate buffers, where pressures on species and ecosystems will be slowed down (*6*), and account for the adaptive genetic variation that can help species to cope with ongoing climatic and landscape alterations (*7*). Likewise, the functional

OUTSIDE THE TOWER

Air quality education in public schools

We began our first lesson on clean air as a virtual meeting in May 2021, with more than 350 fifth to eighth graders and six public school teachers on the screen. Most of the students appeared as black boxes, but as we proceeded, they began to turn on their cameras to ask questions. "If I use scented candles or air fresheners, does that increase air pollution?" one asked. We explained that these examples are indeed sources of particulate matter, which can be harmful if inhaled. Later, we discussed possible pollution mitigation strategies. Although systemic solutions are needed to curb the pollution produced by industry and traffic, carpooling, walking to school, and creating "no idling zones" can all make a difference at the local level. The students were inspired to help and debated the merits of various ways to improve air quality in their schools and homes.

The Clean Air Outreach Project's goal is to increase scientific understanding and promote long-term behavioral changes related to environmental sustainability. We worked with teachers to align our presentation to the curriculum, encouraged them to enhance the atmospheric chemistry content in their science lesson plans, and developed hands-on lab experiments. In the fall of 2021, we brought the program to schools, some of which were located near low-cost air quality multisensor pods installed by our group. To take advantage of the community-specific data, one teacher suggested displaying the air quality dashboard on the televisions in the school hallways. The school hoped to use the information to inform decisions about air quality goals and interventions.

Call for submissions
Outside the Tower is an occasional feature highlighting scientists' advocacy experiences. Submit your advocacy story at http://cts.sciencemag.org.

We were inspired to share our knowledge about air pollution after seeing the global improvement in air quality during the COVID-19 lockdowns. Even though the Waterloo area is ranked the third-fastest-growing region in Canada, it has only one air quality monitoring station. Increased traffic and other sources of pollution disproportionately affect sites that are not well monitored. Because children are vulnerable to the impacts of poor air quality, we wanted to target our outreach to them. We hope that our all-female university-level air quality team of first-generation Canadians will empower these students not only to do what they can to limit air pollution but also to consider following in our footsteps by pursuing careers in science.

Hind A. Al-Abadleh*, Yara Khalaf, Carol Salama, Brenda Kurorwaho
Department of Chemistry and Biochemistry, Wilfrid Laurier University, Waterloo, ON N2L 3C5, Canada.
*Corresponding author. Email: halabadleh@wlu.ca

10.1126/science.abo8050



An air quality multisensor pod installed by the authors can provide data to local schools.

PHOTO: HIND A. AL-ABADLEH

connectivity of protected networks (milestone A.1) should establish climatic corridors that can ensure species' access to analogous future habitats (*8*). These strategies should incorporate measurable indicators to support their effective implementation and monitoring.

Policy-makers face a crucial year to make a meaningful and lasting impact on biodiversity conservation (*9*). The CBD's experts will meet again in June to provide their final recommendations to the post-2020 global biodiversity framework. We call on the experts to advocate for a preventive biodiversity agenda that for once gets ahead of the climate crisis.

André Vicente Liz[1], Duarte Vasconcelos Gonçalves[2], Guillermo Velo-Antón[3], José Carlos Brito[1], Pierre-André Crochet[4], Dennis Rödder[5]*

[1]Research Centre in Biodiversity and Genetic Resources (CIBIO/InBIO), University of Porto, Vairão, Portugal. [2]Interdisciplinary Centre of Marine and Environmental Research (CIIMAR), Matosinhos, Portugal. [3]Departamento de Ecoloxía e Bioloxía Animal, Grupo de Ecoloxía Animal, Torre Cacti (Lab 97), Universidade de Vigo, E-36310 Vigo, Spain. [4]Centre d'Ecologie Fonctionnelle et Evolutive, Centre National de la Recherche Scientifique, Université de Montpellier, École Pratique des Hautes Études, Institut de Recherche pour le Développement, Montpellier, France. [5]LIB, Museum Koenig Bonn, Leibniz Institute for the Analysis of Biodiversity Change, Bonn, Germany.
*Corresponding author.
Email: d.roedder@leibniz-lib.de

REFERENCES AND NOTES

1. A. Arneth *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **49**, 30882 (2020).
2. CBD, "First draft of the post-2020 global biodiversity framework" (2021); www.cbd.int/doc/c/abb5/591f/2e46096d3f0330b08ce87a45/wg2020-03-03-en.pdf.
3. CBD, "Expert input to the post-2020 global biodiversity framework: Transformative actions on all drivers of biodiversity loss are urgently required to achieve the global goals by 2050" (2022); www.cbd.int/doc/c/16b6/e126/9d46160048cfcf74cadcf46d/wg2020-03-inf-11-en.pdf.
4. CBD, "Preparation of the post-2020 global biodiversity framework" (2022); www.cbd.int/doc/c/c949/b2cc/a311c0c411d3a81134e2c7f3/wg2020-03-l-02-en.pdf.
5. C. Román-Palacios, J. J. Wiens, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 4211 (2020).
6. D. Stralberg *et al.*, *Conserv. Lett.* **13**, e12712 (2020).
7. J. O. Hanson *et al.*, *J. Appl. Ecol.* **57**, 2159 (2020).
8. R. A. Senior, J. K. Hill, D. P. Edwards, *Nat. Clim. Change* **9**, 623 (2019).
9. *Nature* **601**, 298 (2022).

10.1126/science.abo7381

TECHNICAL COMMENT ABSTRACTS

Comment on "Discovery of davemaoite, $CaSiO_3$-perovskite, as a mineral from the lower mantle"

Michael J. Walter, Simon C. Kohn, D. Graham Pearson, Steven B. Shirey, Laura Speich, Thomas Stachel, Andrew R. Thomson, Jing Yang

Tschauner *et al*. (Report, 12 November 2021, p. 891) present evidence that diamond GRR-1507 formed in the lower mantle. Instead, the data support a much shallower origin in cold, subcratonic lithospheric mantle. X-ray diffraction data are well matched to phases common in microinclusion-bearing lithospheric diamonds. The calculated bulk inclusion composition is too imprecise to uniquely confirm $CaSiO_3$ stoichiometry and is equally consistent with inclusions observed in other lithospheric diamonds.

Full text: dx.doi.org/10.1126/science.abo0882

Response to Comment on "Discovery of davemaoite, $CaSiO_3$-perovskite, as a mineral from the lower mantle"

Oliver Tschauner, Shichun Huang, Munir Humayun, Wenjun Liu, George R. Rossman

Walter *et al*. issue a number of critical comments on our Report about the discovery of davemaoite to the end that they believe to show that our results do not provide compelling evidence for the presence of davemaoite in the type specimen and that the hosting diamond had formed in the lithosphere. Their claim is based on a misinterpretation of the diffraction data contained in the paper, an insufficient analysis of the compositional data that disregards the three-dimensional distribution of inclusions, and the arbitrary assumption that Earth's mantle shows no lateral variations in temperature, inconsistent with state-of-the-art assessments of mantle temperature variations and with their own published results.

Full text: dx.doi.org/10.1126/science.abo2029

NEXTGEN VOICES: SUBMIT NOW

Shaping the new normal

Add your voice to *Science*! Our new NextGen Voices survey is now open:
Has your institution made any positive changes based on lessons learned during the pandemic? If so, what is the best change and why? If not, what one change do you think should be made based on your own experience?

To submit, go to www.science.org/nextgen-voices
Deadline for submissions is 20 May. A selection of the best responses will be published in the 1 July issue of *Science*. Submissions should be 100 words maximum. Anonymous submissions will not be considered.

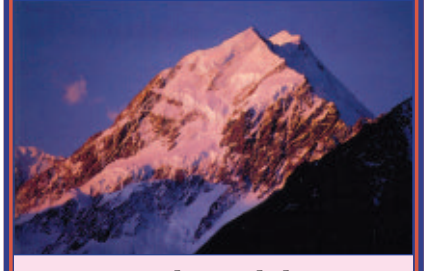
TRAVELS FOR AAAS MEMBERS AND FRIENDS
We invite you to join us!

Discover Egypt
November 6-20, 2022



Discover the amazing antiquities and colossal monuments of Egypt, from the Great Pyramids and Sphinx to the Valley of the Kings, Abu Simbel, Luxor and Aswan.
$4,995 pp + air.

New Zealand!
November 10-25, 2022



Discover the rich heritage of New Zealand with excellent leadership by Lloyd Esler. Explore the geyser fields of Rotorua, Milford Sound, Stewart Island, and some of the world's most spectacular scenery.
$5,495 pp + air.

For a detailed brochure, please call (800) 252-4910
All prices are per person twin share + air

BETCHART EXPEDITIONS INC.
17050 Montebello Rd
Cupertino, California 95014
Email: AAASInfo@betchartexpeditions.com
www.betchartexpeditions.com

RESEARCH

IN *SCIENCE* JOURNALS

Edited by **Michael Funk**



PHOTO: LOURENS SMAK/ALAMY STOCK PHOTO

CLIMATE ECONOMICS

Trade and land-use emissions

International trade allows goods and services produced in one country to be consumed elsewhere, separating consumption from its environmental impacts. Greenhouse gas emissions from land use is one key impact. Hong *et al*. present estimates of land-use emissions embodied in global trade from 2004 to 2017. They found that 27% of all land-use emissions were related to agricultural products consumed in a different country from where they were produced. They also identified the trade relationships representing the largest transfers of land-use emissions, which could help target efforts to improve the sustainability of land use and agricultural production. The authors suggest that improving transparency by regular accounting of land-use emissions embodied in trade could support strategic trade adjustments that would mitigate land-use emissions. —AMS *Science*, abj1572, this issue p. 597

Global shipping encourages land-use change and associated climate emissions to transfer to lower-income countries, even if products are exported.

PALEOECOLOGY

Carbon in the long view

The influence of long-term fluctuations in atmospheric carbon dioxide (CO_2) on woody vegetation growth is complex. The prevailing paradigm has been that it enhances growth. However, using a 500,000-year record of vegetation change at a site in tropical Africa, Gosling *et al*. conclude that shifts in atmospheric CO_2 concentrations do not drive changes in woody cover in the tropics at the millennial scale. Instead, long-term changes in moisture appear to have a greater effect. In the context of current climate change and increasing CO_2 concentrations, these findings suggest, at least for the tropics, that models and policies for anthropogenic climate change mitigation involving enhanced CO_2 drawdown into vegetation may not be as effective as hoped. —AMS

Science, abg4618, this issue p. 653

EVOLUTION

Gene interactions and global fitness

The effect of a particular mutation in a given environment often varies systematically with the fitness of the genetic background in which it arises, but there are different theories as to how such systematic trends manifest. Using a CRISPR gene drive system, Bakerlee *et al.* generated a combinatorial yeast library that allowed them to test the effects of different combinations of missense mutations in 10 genes, including essential ones. Using this system, the authors found that fitness-correlated trends were better explained by an idiosyncratic model, in which they emerged from just a few interactions between specific loci, than they were by a global epistasis model, in which the effects of mutations were generally mediated by background fitness. —MAF

Science, abm4774, this issue p. 630

VACCINES

An Epstein-Barr virus vaccine

Epstein-Barr virus (EBV) infection has been associated with infectious mononucleosis, human cancers, and, most recently, the development of multiple sclerosis. Thus, a vaccine to prevent infection with EBV would be broadly beneficial for human health. Wei *et al.* developed and tested two bivalent EBV vaccines that targeted three or four viral glycoproteins that mediate EBV cell entry into epithelial cells and B cells. The nanoparticle vaccines gH/gL+gp350D_{123} and gH/gL/gp42+gp350D_{123} elicited immune responses in mice, ferrets, and nonhuman primates. Passive transfer of serum from vaccinated mice into humanized mice also conferred protection against EBV, supporting the further development of these bivalent nanoparticle vaccines. —CSM

Sci. Transl. Med. **14**, eabf3685 (2022).

CATALYSIS

Zinc's state in methanol synthesis

Methanol can be synthesized from carbon monoxide (CO), carbon dioxide (CO_2), and

molecular hydrogen (H_2) over copper–zinc (Cu–Zn) catalysts, but studies have disagreed about the chemical state of Zn. Although x-ray photoelectron spectroscopy (XPS) can determine its oxidation state, many studies have been limited to reaction pressures of a few millibar, where the rates are low. Amann *et al*. performed XPS at 180 to 500 millibar for CO_2 and CO hydrogenation over a Zn/ZnO/Cu(211) surface at high turnover rates. Stoichiometric mixtures of CO_2 and H_2 formed ZnO, but for CO and H_2, Zn became more metallic and formed Cu alloys. In industrial synthesis, CO_2 and H_2 are mixed with CO, and the presence of CO would generate Cu–Zn alloy sites active for CO_2 reduction to methanol. —PDS

Science, abj7747, this issue p. 603

HYDROGELS

Bioinspired frequency dampening

Physical activity can interfere with the measurement of physiological signals and often requires band-pass filtering to remove the artifacts. Park *et al*. drew inspiration from the viscoelastic properties of the cuticular pad of spiders that can separate target vibration signals from mechanical noise. The authors developed a gelatin–chitosan hydrogel damper that will absorb frequencies below 30 hertz, whereas higher frequencies transmit due to the temporary disruption of the weak internetwork bonds in the hydrogel. The hydrogel can be used to detect electrophysiological signals such as electroencephalograms and electrocardiograms without interference caused by the patient's walking or breathing. —MSL

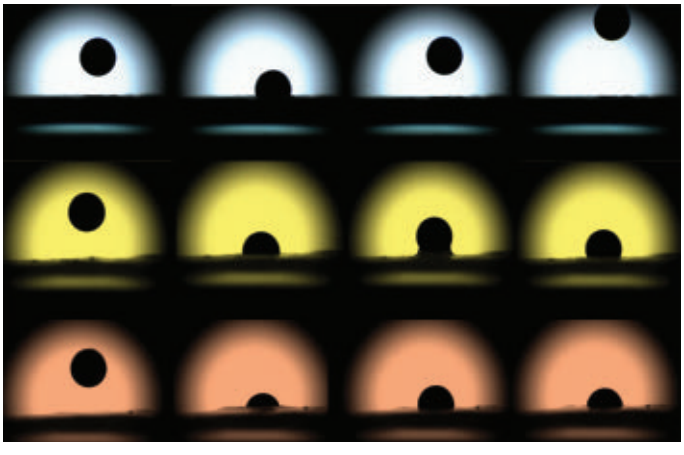
Science, abj9912, this issue p. 624

OPTICS

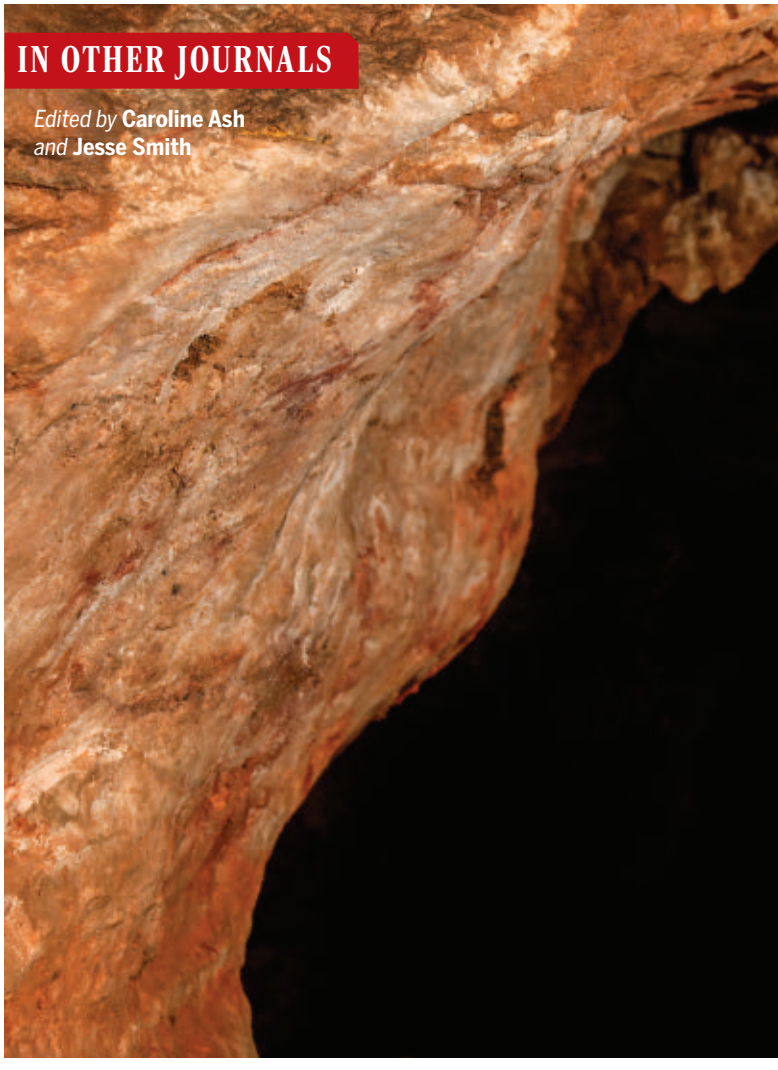
Visualizing motion of quantized vortexes

A quantized vortex is a manifestation of the macroscopic quantum nature of superfluid helium and is essential for understanding its physics. Minowa *et al*. generated silicon nanoparticles in situ in superfluid helium by laser ablation and found that they were attracted to the cores of quantized vortices. By illuminating a light sheet to the silicon nanoparticles, the filaments of the quantum vortices were visualized. Monitoring the motion of two quantized vortex filaments, the authors observed the reconnection process of these filaments. The interaction of light and superfluid helium through silicon nanoparticles could be used to investigate and even manipulate the physical properties of superfluid helium using light in future experiments. —RO

Sci. Adv. 10.1126/sciadv.abn1143 (2022).



A hydrogel coating suppresses the elastic collision of a ball on a hard surface at high temperature (bottom) but less so at cool (middle) or cold (top) conditions.



CELL BIOLOGY

The ER controls microtubule distribution

The endoplasmic reticulum (ER) and microtubules form an intricate, mechanically interdependent network within cells. Although the microtubule cytoskeleton influences the shape of the ER, whether the ER affects the distribution of microtubules is less clear. Tikhomirova *et al*. combined automated quantitative fluorescence microscopy, computer vision, and physics-based modeling and found that changes in ER shape and dynamics had a strong impact on the spatial distribution of microtubules in the cell. Perturbations that affected ER morphology at three-way ER tubule junctions led to the bundling of microtubules. Thus, the ER and microtubules should not be treated as isolated cellular structures because they appear to be highly interdependent, with the ER having a central role in controlling the distribution of microtubules. —SMH

Proc. Natl. Acad. Sci. U.S.A. **119**, e2104309119 (2022).

IMMIGRATION

Deporting criminals, spreading crime

Under the 1996 Illegal Immigration Responsibility Act, the United States substantially increased criminal deportations. Sviatschi shows that the deportations of members of large gangs formed in the United

GRAPHIC: PARK *ET AL.*



FILOVIRUSES

Isolation of Lloviu virus

Filoviruses such as Ebola and Marburg are a perennial source of anxiety for zoonotic transfer. The human populations generally thought to be at risk of filovirus exposure are in sub-Saharan Africa. However, in 2002, the Lloviu filovirus was first identified from a mass die-off of Schreiber's bats *(Miniopterus schreibersii)* in northern Spain. Kemenesi *et al.* report the isolation and sequencing of Lloviu filovirus from bats in Hungary. During 2016–2020, surveillance found Lloviu virus in dead and live cave-dwelling bats and in their blood-engorged ectoparasites. Antibody testing detected nine seropositive bats out of 74 live bats and overall on-site RT-PCR detected four positives among 351 live bats sampled. Although Lloviu virus can infect several human cell lines, unlike Ebola virus, recombinant Lloviu virus does not initiate immune signaling in macrophages and may be nonpathogenic to humans, like the Ebola virus relative Reston filovirus. —CA

Nat. Commun. **13**, 1706 (2022).

Lloviu filovirus has been isolated from Schreiber's bat (*Miniopterus schreibersii*) in Hungary.

PHOTO: JAH/ISTOCK.COM

States by Salvadorans led to a major increase in homicide rates and gang activity such as drug trafficking and extortion in El Salvador. Gang recruitment of children also increased in El Salvador: Children who were exposed to the US-originated gangs were more likely to be involved in gang-related crimes as adults. By increasing gang violence in El Salvador, these deportations fueled an increase in migration of children back to the United States. —BW

Amer. Econ. Rev. https://www.aeaweb.org/articles?id=10.1257/aer.20201540&&from=f (2022).

MOLECULAR METHODS

Regulating the regulators

Human cells contain many messenger RNA (mRNA)–binding proteins (mRBPs) that bind to mRNAs and regulate their translation into functional proteins. Although such mRBPs may outnumber transcription factors, we know much less about how they are regulated by signaling pathways. Vieira-Vieira *et al.* addressed how phosphorylation regulates mRBP function. They developed a method called quantitative RNA interactome capture and used it to measure the fraction of mRBPs that bind to mRNA in relation to phosphorylation. The authors identified individual sites where phosphorylation affects binding to mRNA. Although the data recapitulate known regulatory sites, new potential regulatory sites were also found, including a splicing regulator where phosphorylation affects its nucleocytoplasmic localization. —VV

Mol. Cell 10.1016/j.molcel.2022.03.024 (2022).

GEOPHYSICS

Repeating marsquakes

The single seismometer placed on the surface of Mars by the InSight mission has been in operation for more than 2 years. With a combination of a matched-filter technique and Bedford's law, Sun and Tkalčić report repetitive seismic events associated with the Cerberus Fossae region of Mars. The authors attribute these events to magma movement associated with volcanic activity in upper mantle beneath Cerberus Fossae. If confirmed, these observations would show that the region is seismically active and that the methodology could be useful for detecting other seismic events on the planet. —BG

Nat. Commun. **13**, 1695 (2022).

DEVELOPMENT

The skeleton begins

Most of the vertebrate skeleton solidifies following a softer cartilaginous template that develops from mesenchymal cells that have aggregated into precartilaginous condensations. The shape of the early mesenchymal condensates dictates the shape of the final bone. Paudel *et al.* traced cell lineages through mesenchymal cell condensations as they formed in the developing zebrafish to build the facial skeleton. The locations of mesenchymal condensations are defined by intersecting pro- and anticondensation signals within early developing tissue. The authors suggest that signals from the adjacent epithelium might define positions of the initial mesenchymal condensations, and that maturation to the precartilaginous state might follow as cell proliferation drives daughter cells away from the epithelium. —PJH

Dev. Biol. **486,** 44 (2022).

ENERGY

Electricity price fluctuations in Europe

Both long-term trends and short-term fluctuations in electricity markets can make or break wind and solar producers. Weather patterns have strong correlations with electricity demand while also affecting the amount of electricity that solar and wind farms can supply. Han *et al.* provide a detailed analysis of electricity prices in Europe from 2015 to 2019, including the price fluctuations, averages, and extremes on time scales ranging from days to minutes. This work should help to provide a better understanding of the spot market electricity price time series. —YY

PRX Energy 10.1103/PRXEnergy.1.013002 (2022).

ALSO IN *SCIENCE* JOURNALS

Edited by **Michael Funk**

CORONAVIRUS

Omicron and reinfection risk

So far, our experience with pandemic coronavirus has been that the emergence of new variants is not detected until there has been substantial community transmission. Early in November 2021, South African scientists spotted reinfections consistent with the timing of the emergence of the Omicron variant (B.1.1.529). In a population largely untouched by vaccines but widely infected, Pulliam *et al*. found that Beta or Delta variants rarely caused reinfection. However, after 31 October 2021, individuals were found who had had three experiences of infection (see the Perspective by Zelner and Eisenberg). The culprit was the rapidly emerging Omicron variant, with multiple mutations in the Spike protein. This variant's chief advantage is its ability to evade naturally acquired immunity. Fortunately, the Omicron variant does not fully evade vaccine-derived immunity, but only those privileged to have been vaccinated can benefit. —CA

Science, abn4947, this issue p. 596; see also abp9498, p. 579

PHYSIOLOGY

Understanding heat-induced necrosis

Z-DNA binding protein 1 (ZBP1), a sensor for Z–nucleic acids, is implicated in necrosis caused by heat stress in the worm *Caenorhabditis elegans*. Yuan *et al*. examined the roles of proteins that regulate necrosis in worms undergoing heat stroke. The protein kinase receptor-interacting protein kinase 3 (RIPK3) led to activation of mixed lineage kinase domain-like (MLKL) and Caspase-8, and all were required for heat-induced death. ZBP1 is a RIPK3 and necrosis activator and it appeared to contribute to heat-induced necrosis in the worms. This response did not require the nucleic acid–sensing domain of ZBP1. Transcription of ZBP1 was enhanced by heat shock transcription factor 1 (HSF1). —LBR

Science, abg5251, this issue p.609

CATALYSIS

Peroxide as needed

Hydrogen peroxide for commercial applications is currently mass produced through an indirect process that involves reduction and reoxidation of a quinone. The trouble with making it straight from the elements is that concentrated mixtures of hydrogen and oxygen are dangerously explosive. Lewis *et al*. report that direct peroxide synthesis at safe concentrations is compatible with a current industrial method for producing cyclohexanone oxime, a key precursor to nylon. The protocol could potentially offer substantial efficiency advantages for manufacturing a high-volume bulk chemical. —JSY

Science, abl4822, this issue p. 615

QUANTUM OPTICS

A quantum frequency converter

Practical quantum technologies require interfacing of several subsystems operating in widely different spectral regions. However, most current approaches offer only very small frequency shifts and limited tunability. Tyumenev *et al*. demonstrate efficient (up to 70%) quantum state–preserving photon frequency up-conversion by molecular modulation in hydrogen gas–filled antiresonant photonic crystal fibers (see the Perspective by Sokolov). By combining a high-intensity coherent pump pulse with a (low-intensity) quantum source based on parametric down-conversion (generating pairs of signal and idler photons), the authors show that the idler photons could be up-converted by 125 terahertz using a Raman-scattering process. The process is general and should be applicable to other quantum sources. —ISO

Science, abn1434, this issue p. 621; see also abo2358, p. 575

ICE SHEETS

The deeper story

Shallow, dynamic subglacial water systems provide lubrication that facilitates the movement of overlying ice. But are these thin layers the whole story? Gustafson *et al*. show that the subglacial sediments beneath Whillans Ice Stream in West Antarctica are saturated with a mixture of fossil seawater and freshwater from the glacier (see the Perspective by Chu). This groundwater, extending downward for more than a kilometer, contains more than 10 times as much fluid volume as the shallow hydrologic system above and actively exchanges with it. Therefore, it has the potential to modulate ice streaming and subglacial biogeochemical reactions. —HJS

Science, abm3301, this issue p. 640; see also abo1266, p. 577

CONSERVATION

Population size and risk of extinction

The vaquita porpoise is one of the most endangered animals in the world, with only an estimated 10 individuals remaining. To determine the risk of extinction caused by inbreeding depression, Robinson *et al*. sequenced and examined 20 vaquita genomes to determine their heterozygosity and ancestral population size (see the Perspective by Grueber and Sunnucks). The authors determined that the long-term population size of vaquitas has been low for a marine mammal, with approximately 1000 years of stable genomic diversity. Genomic comparisons with other cetacean species and modeling indicated that vaquitas are unlikely to suffer from inbreeding depression. Therefore, if the risk of bycatch mortality caused by fishermen can be eliminated, then there is a chance that this species will not go extinct. —LMZ

Science, abm1742, this issue p. 635; see also abp9874, p. 574

ECOTOXICOLOGY

Sunscreen turned toxic

Coral reefs face many serious threats from human activity. Sunscreens can cause reef damage, and although the precise mechanisms involved are still under study, some localities have already phased out common components such as oxybenzone. Using a sea anemone as a model system, Vuckovic *et al*. found that oxybenzone is modified within cells by attachment of glucose, turning it from a sunscreen into a potent photosensitizer (see the Perspective by Hansel). The glycoside conjugate is concentrated within the algal symbionts of anemones and corals, and bleached anemones are more susceptible to damage when exposed to ultraviolet light and oxybenzone, suggesting that the algae provide some protection to their hosts. These experiments add to our understanding of reef damage by sunscreens and may help to inform policy and new sunscreen development. —MAF

Science, abn2600, this issue p. 644; see also abo4627, p. 578

FERROELECTRICS

Ultrathin ferroelectric films

The electrical properties of ferroelectrics can be changed with an electric field, making them attractive materials for computer hardware applications. Cheema *et al*. show that extremely thin films of zirconium

dioxide on a silica substrate have ferroelectric order down to the unit cell scale. Whereas in many other materials the ferroelectric behavior is suppressed at the few-nanometer scale, a ferroelectric phase transition occurs if zirconium dioxide is thinner than two nanometers. This property might be true for any fluorite-structured binary oxide, making these types of thin films attractive for next-generation electronics. —BG

Science, abm8642, this issue p. 648

INNATE IMMUNITY

MARCHing against cGAS activation

The binding of microbial or host DNA induces the cytosolic enzyme cyclic GMP-AMP synthase (cGAS) to generate the cyclic nucleotide cyclic GMP-AMP (cGAMP), which leads to the production of type I interferons (IFNs). Yang *et al.* found that ubiquitylation of cGAS by the E3 ubiquitin ligase MARCH8 suppressed the DNA-binding ability of cGAS. This posttranslational modification decreased the production of cGAMP and type I IFN, thereby impairing innate immune responses. Compared with their wild-type counterparts, mice deficient in MARCH8 had enhanced IFN responses to the DNA virus HSV-1. —JFF

Sci. Signal. **15**, eabk3067 (2022).

ANTIBODIES

IgG3 hinge packs a punch

In addition to their extracellular functions in host defense, immunoglobulin G (IgG) antibodies bound to viruses can collaborate with the cytosolic Fc receptor TRIM21 to target and degrade intracellular viruses. Foss *et al.* investigated the role of the elongated and flexible hinge region of human IgG3 in delivering the most potent intracellular killing of adenovirus among the four human IgG subclasses. Use of the IgG3 hinge as a substitute for the hinge region of immunoglobulins from the other IgG subclasses was sufficient to confer enhanced intracellular virus degradation and complement activation dependent on TRIM21. These findings provide fresh insights into how therapeutic monoclonal antibodies can be custom tailored through targeted engineering of the hinge portion of the Fc fragment to boost their antiviral effector activity inside cells. —IRW

Sci. Immunol. **7**, eabj1640 (2022).

CONSERVATION BIOLOGY

Good governance protects protected areas

When habitats degrade, those whose livelihoods depend on them can suffer irreparably. Designating such habitats as protected areas can, at least in theory, reduce degradation while safeguarding livelihoods. In practice, however, this does not always work. Making use of an unprecedented set of long-term data on marine protected and unprotected habitats in Indonesia, Fidler *et al.* uncovered the keys to success. When governance is inclusive, attentive to socioeconomic and cultural conditions, and ensures that vulnerable groups are not marginalized, protected areas work: the proverbial win-win in conservation. —SN

Sci. Adv. 10.1126/sciadv.abl8929 (2022).

MICROBIOLOGY

Understanding bacterial communities

Communities of bacteria exhibit a variety of interactions, including competition between species and cooperation, whereby two species benefit from interacting with each other. These interactions are important for determining the ecology of bacterial communities, so they may lend themselves to manipulations such as engineering a highly competitive bacterial strain that can outcompete a pathogen. In a Perspective, Palmer and Foster discuss the prevalence of cooperation versus competition in various bacterial communities and how work in this area may inform future antibacterial strategies. —GKA

Science, abn5093, this issue p. 581

WETLAND ECOLOGY

Restoring wetlands for carbon

Wetlands disproportionately contribute to carbon sequestration globally. However, the ability of wetlands to store carbon depends on feedbacks between vegetation and geomorphology that allow wetlands to continue to develop over long time periods. When these feedbacks break down, wetlands can become carbon sources. Temmink *et al*. reviewed recent research on the role of plant-landform interactions in wetland carbon storage and the potential for restoration to restore these critical processes. —BEL

Science, abn1479, this issue p. 594

NEUROSCIENCE

The neural basis of cognitive control

What are the neural mechanisms that enable humans to flexibly control and monitor their own actions in various non-routine and novel tasks? Fu *et al*. recorded the activity of more than 1000 neurons in the medial frontal cortex of human epilepsy patients while they performed complex cognitive tasks. They found that domain-general and domain-specific performance-monitoring neurons were intermixed within this brain region. The population activity gave rise to a geometry that allowed domain-general signals to be read out with more than 90% accuracy on single trials while at the same time retaining the ability to separate different conflict conditions. These results show how the human medial frontal cortex resolves the fundamental trade-off between task generalization and specialization, which is critical for cognitive flexibility. —PRS

Science, abm9922, this issue p. 595

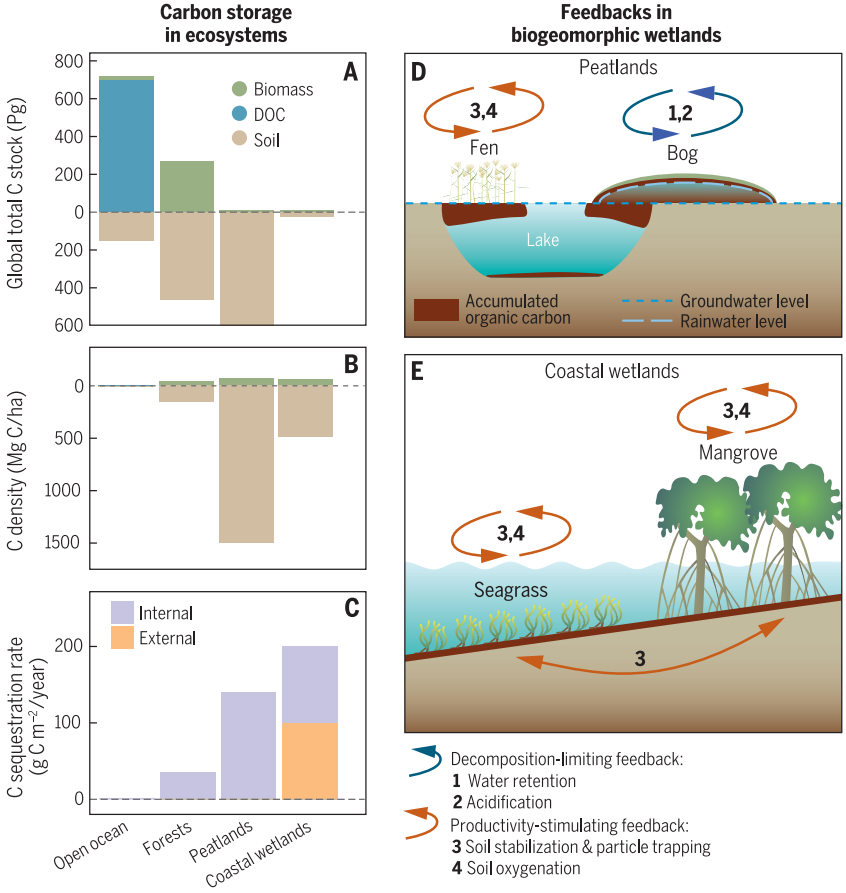
REVIEW SUMMARY

WETLAND ECOLOGY

Recovering wetland biogeomorphic feedbacks to restore the world's biotic carbon hotspots

Ralph J. M. Temmink*, Leon P. M. Lamers, Christine Angelini, Tjeerd J. Bouma, Christian Fritz, Johan van de Koppel, Robin Lexmond, Max Rietkerk, Brian R. Silliman, Hans Joosten, Tjisse van der Heide*

BACKGROUND: Evaluating effects of global warming from rising atmospheric carbon dioxide (CO_2) concentrations requires resolving the processes that drive Earth's carbon stocks and flows. Although biogeomorphic wetlands (peatlands, mangroves, salt marshes, and seagrass meadows) cover only 1% of Earth's surface, they store 20% of the global organic ecosystem carbon. This disproportionate share is fueled by high carbon sequestration rates per unit area and effective storage capacity, which greatly exceed those of oceanic and forest ecosystems. We highlight that feedbacks between geomorphology and landscape-building wetland vegetation underlie these critical qualities and that disruption of these biogeomorphic feedbacks can switch these systems from carbon sinks into sources.



Carbon storage in biogeomorphic wetlands. Organic carbon (**A**) stocks, (**B**) densities, and (**C**) sequestration rates in the world's major carbon-storing ecosystems. Oceans hold the largest stock, peatlands (boreal, temperate, and tropical aggregated) store the largest amount per unit area, and coastal ecosystems (mangroves, salt marshes, and seagrasses aggregated) support the highest sequestration rates. (**D** and **E**) Biogeomorphic feedbacks, indicated with arrows, can be classified as productivity stimulating or decomposition limiting. Productivity-stimulating feedbacks increase resource availability and thus stimulate vegetation growth and organic matter production. Although production is lower in wetlands with decomposition-limiting feedbacks, decomposition is more strongly limited, resulting in net accumulation of organic matter. (D) In fens, organic matter accumulation from vascular plants is amplified by productivity-stimulating feedbacks. Once the peat rises above the groundwater and is large enough to remain waterlogged by retaining rainwater, the resulting bog maintains being waterlogged and acidic, resulting in strong decomposition-limiting feedbacks. (E) Vegetated coastal ecosystems generate productivity-stimulating feedbacks that enhance local production and trapping of external organic matter.

ADVANCES: A key advancement in understanding wetland functioning has been the recognition of the role of reciprocal organism-landform interactions, "biogeomorphic feedbacks." Biogeomorphic feedbacks entail self-reinforcing interactions between biota and geomorphology, by which organisms—often vegetation—engineer landforms to their own benefit following a positive density-dependent relationship. Vegetation that dominates major carbon-storing wetlands generate self-facilitating feedbacks that shape the landscape and amplify carbon sequestration and storage. As a result, per unit area, wetland carbon stocks and sequestration rates greatly exceed those of terrestrial forests and oceans, ecosystems that worldwide harbor large stocks because of their large areal extent.

Worldwide biogeomorphic wetlands experience human-induced average annual loss rates of around 1%. We estimate that associated carbon losses amount to 0.5 Pg C per year, levels that are equivalent to 5% of the estimated overall anthropogenic carbon emissions. Because carbon emissions from degraded wetlands are often sustained for centuries until all organic matter has been decomposed, conserving and restoring biogeomorphic wetlands must be part of global climate solutions.

OUTLOOK: Our work highlights that biogeomorphic wetlands serve as the world's biotic carbon hotspots, and that conservation and restoration of these hotspots offer an attractive contribution to mitigate global warming. Recent scientific findings show that restoration methods aimed at reestablishing biogeomorphic feedbacks can greatly increase establishment success and restoration yields, paving the way for large-scale restoration actions. Therefore, we argue that implementing such measures can facilitate humanity in its pursuit of targets set by the Paris Agreement and the United Nations Decade on Ecosystem Restoration. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: r.j.m.temmink@uu.nl (R.J.M.T.); tjisse.van.der.heide@nioz.nl (T.v.d.H)
Cite this article as R. J. M. Temmink *et al.*, *Science* **376**, eabn1479 (2022). DOI: 10.1126/science.abn1479

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abn1479

REVIEW

WETLAND ECOLOGY

Recovering wetland biogeomorphic feedbacks to restore the world's biotic carbon hotspots

Ralph J. M. Temmink[1,2,3]*, Leon P. M. Lamers[3,4], Christine Angelini[5], Tjeerd J. Bouma[6,7,8,9], Christian Fritz[3,10], Johan van de Koppel[6,7], Robin Lexmond[11], Max Rietkerk[1], Brian R. Silliman[12], Hans Joosten[13], Tjisse van der Heide[2,7]*

Biogeomorphic wetlands cover 1% of Earth's surface but store 20% of ecosystem organic carbon. This disproportional share is fueled by high carbon sequestration rates and effective storage in peatlands, mangroves, salt marshes, and seagrass meadows, which greatly exceed those of oceanic and forest ecosystems. Here, we review how feedbacks between geomorphology and landscape-building vegetation underlie these qualities and how feedback disruption can switch wetlands from carbon sinks into sources. Currently, human activities are driving rapid declines in the area of major carbon-storing wetlands (1% annually). Our findings highlight the urgency to stop through conservation ongoing losses and to reestablish landscape-forming feedbacks through restoration innovations that recover the role of biogeomorphic wetlands as the world's biotic carbon hotspots.

Global warming, resulting from rapidly rising atmospheric carbon dioxide (CO_2) concentrations since the Industrial Revolution, has increasingly drawn attention toward understanding and quantifying the processes that drive Earth's carbon stocks and flows (*1*, *2*). Burial of organic matter remains the largest carbon sequestering process on the planet, rivaled only by the ocean's inorganic carbon solubility pump (*3*, *4*). Although wetlands cover just 2% of Earth's surface (*5*), they store more than 20% of global organic ecosystem carbon (all live and dead organic matter from terrestrial, freshwater, and oceanic systems combined) (*4*, *6*). Moreover, wetland carbon sequestration rates can be orders of magnitude higher as compared with those of terrestrial and oceanic ecosystems (*7*). Recent work has addressed the importance of wetlands as natural climate solutions and the cost-effectiveness of their restoration (*8*, *9*). However, restoring carbon storage functions requires an understanding of the mechanisms that underlie their large carbon stocks and high sequestration rates.

An important advancement in understanding wetland functioning has been the recognition of the key role of reciprocal organism-landform interactions: so-called biogeomorphic feedbacks (*10*, *11*). Biogeomorphic feedbacks entail self-reinforcing interactions between biota and geomorphology, by which organisms—often vegetation—engineer landforms through positive density-dependent relationships. Here, we focus on the major wetlands that are shaped by such vegetation-geomorphology feedbacks: (i) peatlands, where vegetation retains water by preventing lateral and vertical seepage, yielding landforms shaped by vertical and horizontal peat accretion (*12*), and (ii) coastal wetlands—including seagrass meadows (*13*), salt marshes (*10*), and mangroves (*14*)—where vegetation traps suspended particles from the water and stabilizes underlying soils to form elevated landscape features. Although it has been known for two centuries that vegetation-driven feedbacks shape "biogeomorphic wetlands" (*15*), the role of these feedbacks in controlling carbon sequestration and storage have received insufficient attention.

In this Review, we first compare the carbon stocks and sequestration rates of the three major carbon-storing ecosystems—oceans, forest, and wetlands—after which we highlight how vegetation-geomorphology feedbacks shape wetland landscapes and their role as global carbon hotspots. We summarize how anthropogenic disruption transforms these carbon sinks and stocks into sources and highlight how implementing new restoration designs aimed at jumpstarting and sustaining biogeomorphic feedbacks may improve carbon sequestration.

Comparing organic carbon stocks and sequestration rates between ecosystems

Our literature-based compilation highlights that the major carbon-storing wetlands store the bulk of their organic carbon as soil organic matter, whereas oceans and forests hold most of their carbon in the water layer and living biomass, respectively (Fig. 1A) (*16*). Although oceans and forests hold massive amounts of organic carbon because of their large spatial extent, their area-specific carbon density (carbon stock per unit area) is smaller compared with that of biogeomorphic wetlands (Fig. 1B). Carbon density is highest in peatlands (1000 to 2000 Mg C ha^{-1}), followed by mangroves (900 Mg C ha^{-1}), salt marshes (400 Mg C ha^{-1}), and seagrass meadows (330 Mg C ha^{-1}). Carbon density is lower in terrestrial forests (150 to 230 Mg C ha^{-1}) and much lower in the oceans (2.4 Mg C ha^{-1}) (*17*, *18*).

Recent sequestration rates of internally and externally produced organic carbon per unit area over the past 10 to 125 years are higher in tropical peatlands (200 g C m^{-2} $year^{-1}$) as compared with their boreal (100 g C m^{-2} $year^{-1}$) and temperate (120 g C m^{-2} $year^{-1}$) counterparts (Fig. 1C). Average salt marsh and mangrove sequestration rates (250 and 200 g C m^{-2} $year^{-1}$, respectively) may outpace or equal those of tropical peatlands, and seagrass meadows bury 150 g C m^{-2} $year^{-1}$, which is more than boreal and temperate but less than tropical peatlands (*7*, *19*). For coastal ecosystems, 100 g C m^{-2} $year^{-1}$ originates from external (such as riverine and marine) sources, which gets trapped and buried (*20*–*23*). All of these vegetated wetland rates are higher than those of terrestrial forests and oceans, where net sequestration rates are below 50 g C m^{-2} $year^{-1}$ (Fig. 1C). Intact vegetated coastal wetlands and freshwater peatlands worldwide currently sequester 0.7 Pg C $year^{-1}$, equaling 6% of the total annual global anthropogenic carbon emissions (which were estimated in 2019 to be 11.5 Pg C) (*4*).

Biogeomorphic feedbacks shape wetland carbon storage hotspots

In 45% of all wetlands worldwide, biogeomorphic feedbacks shape landscape formation

[1]Environmental Sciences, Copernicus Institute of Sustainable Development, Utrecht University, Princetonlaan 8a, 3584 CB, Utrecht, Netherlands. [2]Department of Coastal Systems, Royal Netherlands Institute for Sea Research, 1790 AB Den Burg, Netherlands. [3]Aquatic Ecology and Environmental Biology, Radboud Institute for Biological and Environmental Sciences, Radboud University, Heyendaalseweg 135, 6525 AJ Nijmegen, Netherlands. [4]B-WARE Research Centre, Toernooiveld 1, 6525 ED Nijmegen, Netherlands. [5]Department of Environmental Engineering Sciences, Engineering School for Sustainable Infrastructure and Environment, University of Florida, Post Office Box 116580, Gainesville, FL 32611, USA. [6]Department of Estuarine and Delta Systems, Royal Netherlands Institute for Sea Research, 4401 NT Yerseke, Netherlands. [7]Conservation Ecology Group, Groningen Institute for Evolutionary Life Sciences, University of Groningen, 9700 CC Groningen, Netherlands. [8]Building with Nature group, HZ University of Applied Sciences, Postbus 364, 4380 AJ Vlissingen, Netherlands. [9]Faculty of Geosciences, Department of Physical Geography, Utrecht University, 3508 TC Utrecht, Netherlands. [10]Integrated Research on Energy, Environment and Society (IREES), University of Groningen, Nijenborgh 6, Groningen, 9747 AG, Netherlands. [11]Experimental Plant Ecology, Radboud Institute for Biological and Environmental Sciences, Radboud University, Heyendaalseweg 135, 6525 AJ Nijmegen, Netherlands. [12]Division of Marine Science and Conservation, Nicholas School of the Environment, Duke University, 135 Duke Marine Lab Road, Beaufort, NC, USA. [13]Institute of Botany and Landscape Ecology, Greifswald University, Partner in the Greifswald Mire Centre, Soldmannstrasse 15, 17487 Greifswald, Germany.
*Corresponding author. Email: r.j.m.temmink@uu.nl (R.J.M.T.); tjisse.van.der.heide@nioz.nl (T.v.d.H)

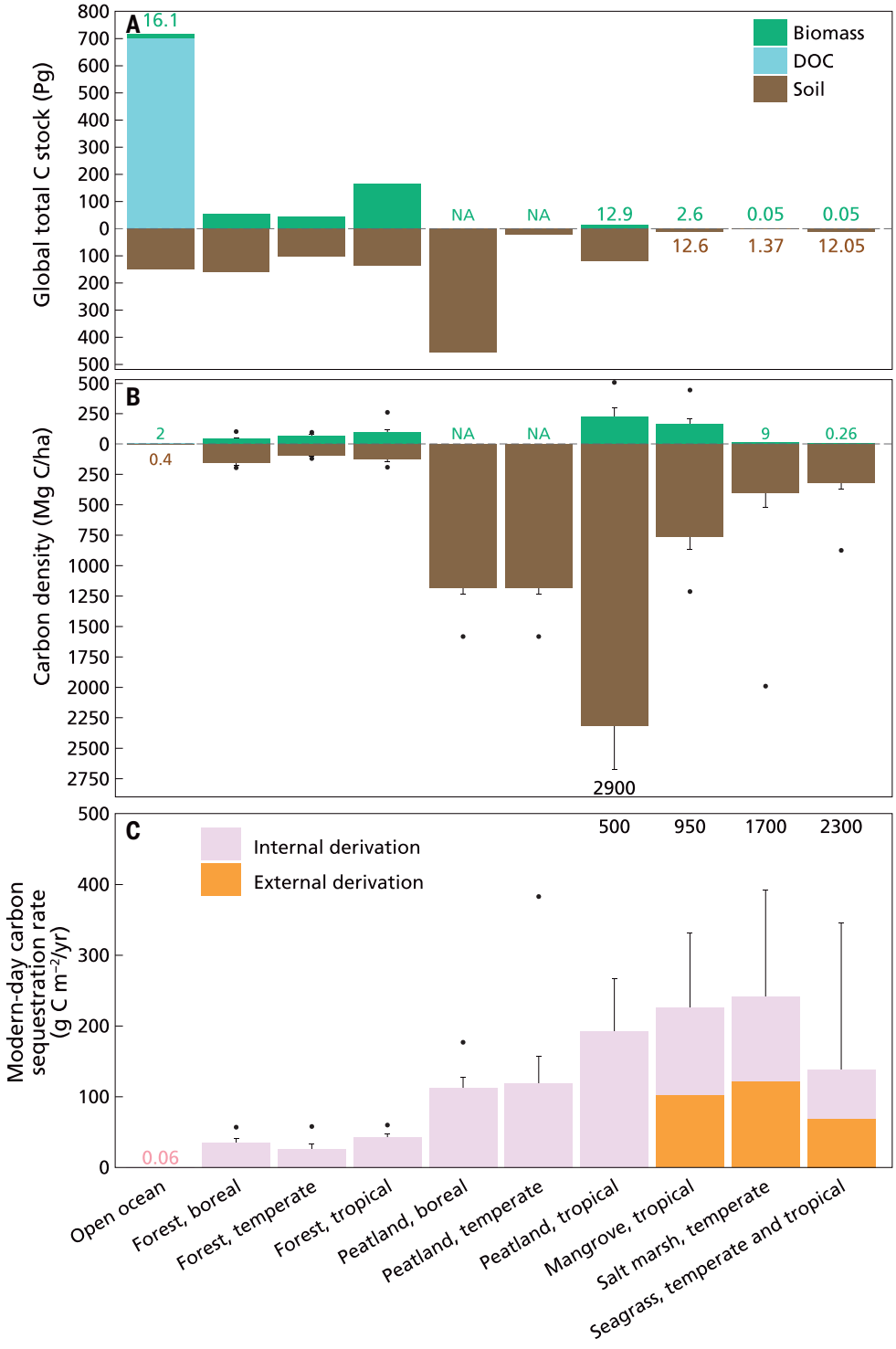


Fig. 1. Overview of the world's major carbon-storing ecosystems. (**A**) Global total carbon stocks. (**B**) Carbon densities. (**C**) Modern-day carbon sequestration rates. Oceans hold the largest stocks globally in the form of dissolved organic carbon (DOC; >97% of the carbon pool), whereas peatlands store the largest amounts of carbon per unit area. Coastal ecosystems generate the highest modern-day sequestration rates (mean rate over the past 10 to 125 years) by storing both locally and externally derived organic matter. This process, however, may become self-limiting when sediment elevation outpaces sea level rise, which is unlikely under current climate change. In addition, damage from stochastic disturbances such as storm-induced erosion can also limit long-term storage. Error bars in (B) and (C) indicate SD; black dots (or numbers when they fall outside the *y* axis) depict observed maxima. We could not calculate uncertainties for the ocean because these values were calculated from global estimates (*16*). Data were generally collected from recent synthesis studies per ecosystem type (*16*). Carbon sequestration rates are from periods ranging from 10 to 125 years (recent apparent rate of carbon accumulation, which are higher than long-term rates over ~10,000 years because of continued decomposition of accumulated matter) (*36*). References and methodological details are provided in table S1 (*16*). [Figure design: Ton A. W. Markus]

and carbon capture and storage processes (Fig. 2 and Table 1) (*5*). Two overarching types of feedbacks control the ability of biogeomorphic wetlands to capture and store carbon. In wetlands driven by "productivity-stimulating" feedbacks, landscape formation and carbon storage are enhanced by feedback processes that increase resource availability and thus stimulate vegetation growth and organic matter production. In wetlands shaped by "decomposition-limiting" feedbacks and consequent nutrient immobilization, production is slower, but because decomposition is more strongly limited, organic matter can accumulate in such wetland soils (Fig. 2).

Peatlands

Peatlands are effective organic carbon sinks in terms of long-term storage per unit area (Fig. 1B). Peatland landforms are shaped by landscape-scale interactions between plants, peat, and water (*12*). Their formation is typically initiated through one of two processes: (i) paludification and (ii) terrestrialization. Paludification is the process in which a change in the hydrological balance shifts a previously drier, vegetated, and inorganic soil terrestrial ecosystem toward a peat-accumulating, biogeomorphic wetland ecosystem (*24*). By contrast, terrestrialization occurs in aquatic

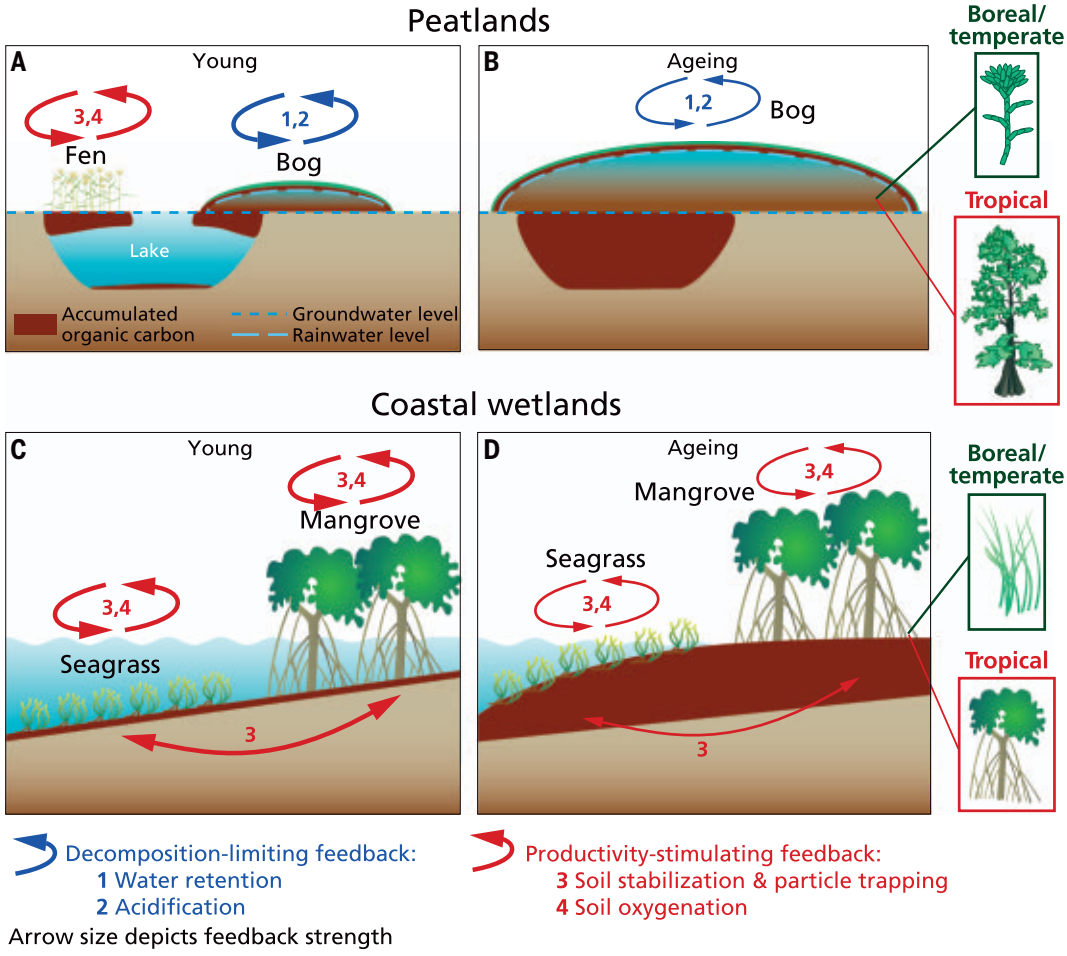


Fig. 2. Conceptual representation of the formation of carbon-storing biogeomorphic wetlands. Density-dependent processes underlying biogeomorphic feedbacks can be classified as productivity-stimulating or decomposition-limiting. (**A**) Peatland formation is initiated through either terrestrialization or paludification. Terrestrialization of aquatic systems by accumulation of organic matter from vascular plants is amplified by productivity-stimulating feedbacks in fens, whereas paludification initiates directly over mineral soil. (**B**) Once the peat surface rises above the groundwater level, and the peat is large enough to remain waterlogged by retaining rainwater, the resulting bog maintains waterlogged and acidic conditions, resulting in strong decomposition-limiting feedbacks. (**C**) Vegetated coastal ecosystems (seagrass meadows, mangroves, and salt marshes) generate productivity-stimulating feedbacks that stimulate local production and substrate building. (**D**) This process can become self-limiting as the system ages because increasing sediment elevation limits further development when this process outpaces sea level rise. This is not a comprehensive representation of all feedbacks. Boxes with vegetation indicate dominant vegetation type in boreal or temperate and tropical wetlands, respectively. [Figure design: Ton A. W. Markus. Symbols are from Integration and Application Network, IAN Image Library (https://ian.umces.edu/imagelibrary)]

Table 1. Global extent (million hectares) of (near) natural biogeomorphic wetlands, lost or degraded (%) and the annual rate of human-induced losses (% year^{-1}). The range shows minimum to maximum and the central value (square brackets).

Ecosystem	Climate zone	Global extent (million hectare)	Lost or degraded (%)	Annual loss rate (% year^{-1})	Global extent data source	Loss/degraded data source	Annual loss rate data source
Peatland	Boreal*	386 (170†)	4% (15%†)	0% (0.9%†)	(60)	(60)	(90)
Peatland	Temperate	19	57%	0%‡	(60)	(60)	(90)
Peatland	Tropical	59	41%	3.3%	(60)	(60)	(91)
Mangrove	Tropical	17	35%	0.7 to 3.0 [1.9]%	(92)	(63)	(74)
Salt marsh	Temperate	6	42%	1.0 to 2.0 [1.5]%	(93)	(61)	(74)
Seagrass meadow	Temperate and tropical	18 to 60 [39]	29%	0.4 to 2.6 [1.5]%	(94)	(62)	(74)

*Data includes polar and boreal peatlands. †Circa half of the boreal peatlands can be classified as permafrost peatlands (68). Their preindustrial extent was ~200 million ha, but because of human-induced climate warming, 15% of permafrost peatlands have been degraded at a loss rate of 0.9% since 1850, currently leaving 170 million ha. ‡Temperate peatlands are slowly increasing in extent owing to rewetting and restoration of degraded or drained peatlands (~300,000 ha in total).

systems, such as shallow lakes, when organic matter deposition outpaces its decomposition in the anoxic environment, resulting in the gradual in-filling of the water basin over time (25). As the waterbody accretes organic matter, it transitions into a "fen," a peatland under the influence of ground or surface water and often dominated by emergent fast-growing vascular plants such as grasses, rushes, and sedges (25). In fens, plant growth is supported by a productivity-stimulating positive feedback in which the vegetation's root mat traps and fixes the produced dead organic matter and maintains effective water storage through large pores and surface oscillation (26–28). Supported by oxygen released from the root mat, the labile organic matter decomposes rapidly, releasing nutrients that in turn stimulate plant

growth (29, 30). The more recalcitrant fraction with the highest carbon percentage, however, remains and accumulates (31).

Once the peat surface rises above the groundwater, the system transitions into a bog in which decomposition-limiting feedbacks facilitate landscape formation (12). Bogs are fed primarily by rainwater, which is retained within the landform by both the vegetation—*Sphagnum* moss in cool region bogs, and trees in the tropics—and the accumulated peat layer (12). The plants, and their detrital remains, limit lateral and vertical drainage and regulate evaporation. As a result, soils remain persistently waterlogged, acidic, anoxic, and nutrient-poor; these conditions hamper the establishment of competitive species and stifle organic matter decomposition (12, 32–34).

The self-reinforcing biogeomorphic feedback between vegetation development, water retention, and peat accumulation yields a biogenic landscape that forms over a period of hundreds to thousands of years, with long-term peat and carbon accumulation rates of 1 to 3 mm $year^{-1}$ and on average 18 g C m^{-2} $year^{-1}$ [which is lower than modern sequestration rates due to continued decomposition (Fig. 1C)] (35, 36). Primary production is higher in tropical peatlands than boreal and temperate ones and is quantitatively different because of the production of lignin (37, 38), which allows for higher sequestration rates (Fig. 1C).

Coastal wetlands

Compared with peatlands, seagrass meadows, salt marshes, and mangrove forests are generally more productive and are driven by productivity-stimulating feedbacks (38, 39). Whereas peatlands generally have low inputs of external organic C, coastal wetlands commonly receive organic matter from the ocean and from rivers and thus sequester both externally and locally produced organic matter (20, 40). By attenuating currents and waves with their aboveground vegetation structures, coastal wetlands can trap large amounts of externally produced, suspended organic particles that end up buried in the root-stabilized anoxic soils (13, 41). The ratio of locally versus externally produced organic matter differs widely depending on wetland size, vegetation, and location (20, 42), with close proximity to productive coastal waters or rivers favoring allochthonous input (43, 44). Moreover, large wetlands with dense and stiff vegetation also tend to dissipate more hydrodynamic energy, favoring entrapment of incoming particles (45, 46). Externally produced organic material often appears to be much more recalcitrant than the internally produced fraction (47). This highlights that the filtering function of these wetlands may rival their local productivity in importance for carbon sequestration, because on average, almost 50% of all buried organic carbon originates from external sources, although this value varies with context (20–23) (Fig. 1C).

Regardless of its origin, the presence of organic matter in vegetated coastal wetlands creates a productivity-stimulating positive feedback. Decomposition of labile organic matter fueled by radial oxygen loss from plant roots (48) stimulates in situ plant production, while the more recalcitrant fraction is stored in the sediment layers (40, 49). In addition, soil stabilization and attenuation of hydrodynamic forces reduce losses from uprooting and erosion during storms, while the active trapping of particles from the water column also increases water clarity (13, 50), enhancing underwater light availability and favoring the growth of seagrass meadows (13). In salt marshes and mangroves, the trapping of particles increases the bed level, reducing inundation stress (51). Moreover, reciprocal facilitation between coastal vegetation and associated biota can further amplify carbon storage (52, 53). Last, an increasing number of studies highlight the importance of landscape-scale reciprocal interactions between coastal ecosystems. Specifically, seagrasses have been found to facilitate marsh and mangrove establishment through their attenuation of waves (54), and marshes and mangroves trap suspended particles to improve water clarify and facilitate adjacent seagrasses.

These multiple—and in many cases, cross-ecosystem—productivity-stimulating biogeomorphic feedbacks result in highly productive wetland complexes, with soils that rapidly accrete, both vertically and laterally, over time in the initial phase of development (55). In salt marshes, sediment accretion rates can reach up to 25 mm $year^{-1}$, whereas in mangroves and seagrasses, rates can be as high as 21 and 10 mm $year^{-1}$, respectively (56). As these ecosystems age and develop, their sediment accumulation rates may keep pace with or even exceed sea level rise (current relative sea level rise, 0 to 10 mm $year^{-1}$) (57, 58). When sediment accretion rates exceed relative sea level rise, local carbon accumulation levels out as the increasing surface elevation decreases water saturation (higher decomposition) and flooding frequency (lower organic matter import) (57, 59).

Human-induced breakdown of feedbacks: From carbon sink to source

Many biogeomorphic wetlands have been rapidly deteriorating and continue to decline in area at rates that range from 0.4 to 3.3% per year, with the exception of cooler-region, boreal peatlands that have remained stable (Table 1). Salt marshes have declined by 42%, and mangroves and seagrass meadows have lost 35 and 29% of their area over the past centuries, respectively (60–63). These losses are caused by habitat destruction from land-use change, overexploitation, eutrophication, salinization, trophic cascades, and climate change–related extreme events such as heat waves and increased storm magnitude and frequency (64, 65). In the future, sea level rise will likely result in major loss of coastal wetlands and their carbon stocks, particularly in areas where landward migration is hampered by human infrastructure—a phenomenon called "coastal squeeze" (66). Temperate and tropical peatlands have been degraded by 57 and 41% in their areal extent, respectively, mostly because of land-use changes, exploitation, and wildfires (60, 67). By contrast, boreal peatlands have not been rapidly declining in their overall extent (<5% loss). However, climate change–driven thawing of the permafrost, which encompasses about half of all boreal peatlands, has affected 15% of these coldest peatlands. The net effect of permafrost thaw on the climate remains unknown because permafrost thaw increases methane (CH_4) and CO_2 emissions from increased decomposition rates while simultaneously increasing productivity and carbon sequestration (68, 69).

At present, biogeomorphic wetlands worldwide experience average annual loss rates of around 1%, with associated yearly carbon losses amounting to 0.5 Pg C (Table 1), which would account for 5% of the current anthropogenic carbon emissions (11.5 Pg C) (4). In contrast to the immediate carbon losses from logging of forests, land-use changes in biogeomorphic wetlands do not necessarily result in the immediate removal of most carbon because the bulk of the carbon is stored in the soil (Fig. 1). Specifically, conversion of peatlands to agricultural land results in instant carbon loss owing to the removal of any aboveground biomass (70), but this is followed by a continued loss of soil organic carbon in the following century (Fig. 3) (71, 72). Loss of coastal wetland vegetation commonly results in rapid erosion and oxidation of carbon-rich soils because the vegetation no longer stabilizes the soil (73, 74). However, in regions where coastal wetlands are "reclaimed" under the protection of levees or dikes, erosion from currents and waves is obviously unimportant, causing accumulated organic matter to oxidize much more gradually (61).

Conservation and restoration of carbon hotspots

Our findings emphasize the importance of conserving and restoring biogeomorphic wetlands worldwide. Conservation measures are particularly rewarding in peatlands, where carbon densities are the highest and where carbon stocks lost by degradation take centuries to millennia to rebuild. Complementary to conservation, restoration of degraded biogeomorphic wetlands and their carbon storage and sink function should be a key element of our global carbon strategy. Restoration is likely to

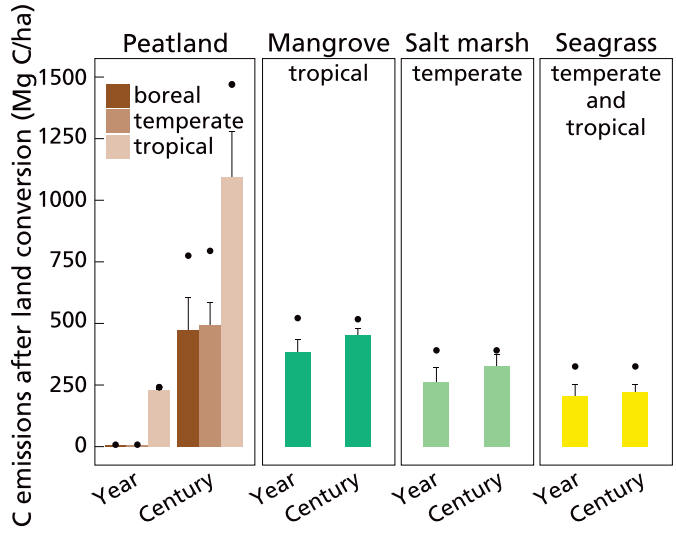


Fig. 3. Carbon emissions after land-use change in biogeomorphic wetlands. Land-use change and (subsequent) chemical and physical erosion result in rapid carbon losses in coastal systems. "Year" indicates 1 year loss. Although carbon losses in peatlands can also be high upon land-use change (for example, logging of tropical forests), they are typically lower but continue for centuries at a slower pace, resulting in higher overall carbon losses. "Century" indicates loss over 100 years. Error bars indicate SD; black dots indicate observed maxima. We assumed instantaneous emissions from biomass after land conversion. For coastal systems, loss of carbon after land conversion was assumed 25 to 100% after year 1 and 63 to 100% after 100 years (*74*), whereas for peatlands, we applied commonly used land-use emission factors to calculate long-term losses (*60*, *72*). References and methodological details are provided in table S2 (*16*). [Figure design: Ton A. W. Markus]

be most rewarding over shorter time scales in both high–carbon stock systems (where emissions can be avoided) and high-productivity systems (where fast sequestration takes place). Coastal wetlands can offer great potential for fast carbon accumulation by sequestering both externally and internally produced material on a time scale of years to decades (*75*). Although carbon sequestration rates of peatlands are slower than those in coastal systems, achieved gains from restoration can still be high because these measures reduce currently ongoing large emissions from these areas (*72*).

Because of the benefits for carbon storage and other ecosystem services, conservation practitioners and policy-makers increasingly consider restoration of biogeomorphic wetlands as a viable tool to counteract mounting losses (*76*, *77*). At present, however, restoration of these systems is often ineffective (generally <50% success) (*76*) and costly compared with restoration of other ecosystem types. For example, restoration costs of terrestrial ecosystems such as grasslands, woodlands, and temperate and tropical forests range from 500 to 5000 US $/ha (*77*), with restoration scales ranging from <1000 to >100,000 ha (*78*). By contrast, restoration of vegetated biogeomorphic wetlands most often occurs at spatial scales of 0.1 ha to tens of thousands of hectares, with costs ranging from 750 to 1,000,000 US$/ha (*76*, *79*). An important issue underlying these low success rates and high costs is that biogeomorphic feedbacks only work beyond a certain minimum vegetation patch size and density (*80*). Below these thresholds, unpredictable losses occur, while natural establishment is hampered (*13*, *81*). In such cases, a so-called "Window of Opportunity" may be required—a rare period of conditions that are particularly beneficial for vegetation establishment and allow vegetation to grow beyond the size or density threshold required for the biogeomorphic feedback to initiate and support longer-term survival (*82*).

Despite the importance of facilitation by biogeomorphic feedbacks in wetlands, classic restoration approaches have been strongly influenced by agriculture and forestry science, which typically plant in dispersed spatial configuration with the aim of minimizing competition (*83*). Recent advancements now emphasize the importance of facilitation over competition in these systems. In coastal wetlands, restoration experiments demonstrate that large-scale approaches favor facilitative interactions and are therefore typically more successful (*84*). Similarly, facilitation can be harnessed at smaller scales by planting in clumps rather than applying plantation-style dispersed designs, a change that was found to double restoration yields (*83*). Moreover, the same can be achieved when individual small seagrass or marsh grass plants are transplanted within biodegradable structures that temporarily mimic facilitating effects of larger patches, such as suppression of waves and sediment mobility (*46*, *85*). Last, depending on the system, it may also be possible to artificially create a Window of Opportunity with engineering measures to allow natural reestablishment (*86*).

Similar to coastal wetlands, peatland restoration has been most successful when recovering natural conditions through large, landscape-scale rewetting measures. This is particularly the case for peat bogs, where inserting dams to restore water retention in degraded bogs has been successful because it creates a window of opportunity for natural plant-hydrology feedbacks to reestablish (*87*). Sphagnum paludiculture, a new form of peat bog culturing, takes this approach one step further; after rewetting, peatmosses are actively introduced at a sufficient spatial scale to overcome establishment thresholds and allow their sustainable harvest (*88*). Similarly, paludiculture in fens focuses on large-scale reintroduction and sustainable harvest of rapidly growing helophytes, such as *Typha* sp., thus reestablishing productivity-stimulating feedbacks (*88*). Last, recent work revealed that peatland rewetting strategies in general can be improved by striking the best balance between stopping sustained CO_2 emissions from drainage and CH_4 release from rewetting by optimizing the water table height (*72*, *89*).

On the basis of this synthesis, we argue that stopping biogeomorphic wetland losses through conservation measures is of utmost importance. Moreover, recent technical advancements that focus on recovery of landscape-forming feedbacks have now paved the way for large-scale restoration that reverts biogeomorphic wetlands from sources back to sinks. Therefore, we argue that implementation of conservation measures combined with restoration actions can enhance the role of biogeomorphic wetlands as natural climate solutions, facilitating humanity to reach the targets set by the Paris Agreement and the United Nations Decade on Ecosystem Restoration.

REFERENCES AND NOTES

1. H. D. Matthews, N. P. Gillett, P. A. Stott, K. Zickfeld, The proportionality of global warming to cumulative carbon emissions. *Nature* **459**, 829–832 (2009). doi: 10.1038/nature08047; pmid: 19516338
2. M. R. Allen *et al.*, Warming caused by cumulative carbon emissions towards the trillionth tonne. *Nature* **458**, 1163–1166 (2009). doi: 10.1038/nature08019; pmid: 19407800
3. P. Ciais *et al.*, The physical science basis. Contribution of working group I to the fifth assessment report of the intergovernmental panel on climate change. *Chang. IPCC Clim.* **18**, 465–570 (2013).
4. P. Friedlingstein *et al.*, Global carbon budget 2020. *Earth Syst. Sci. Data* **12**, 3269–3340 (2020). doi: 10.5194/essd-12-3269-2020
5. N. C. Davidson, E. Fluet-Chouinard, C. M. Finlayson, Global extent and distribution of wetlands: Trends and issues. *Mar. Freshw. Res.* **69**, 620–627 (2018). doi: 10.1071/MF17019
6. Y. M. Bar-On, R. Phillips, R. Milo, The biomass distribution on Earth. *Proc. Natl. Acad. Sci. U.S.A.* **115**, 6506–6511 (2018). doi: 10.1073/pnas.1711842115; pmid: 29784790
7. E. McLeod *et al.*, A blueprint for blue carbon: Toward an improved understanding of the role of vegetated coastal habitats in sequestering CO2. *Front. Ecol. Environ.* **9**, 552–560 (2011). doi: 10.1890/110004

8. B. W. Griscom *et al*., Natural climate solutions. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 11645–11650 (2017). doi: 10.1073/pnas.1710465114; pmid: 29078344
9. P. Taillardat, B. S. Thompson, M. Garneau, K. Trottier, D. A. Friess, Climate change mitigation potential of wetlands and the cost-effectiveness of their restoration. *Interface Focus* **10**, 20190129 (2020). doi: 10.1098/rsfs.2019.0129; pmid: 32832065
10. C. Schwarz *et al*., Self-organization of a biogeomorphic landscape controlled by plant life-history traits. *Nat. Geosci.* **11**, 672–677 (2018). doi: 10.1038/s41561-018-0180-y
11. C. Wang, S. Temmerman, Does biogeomorphic feedback lead to abrupt shifts between alternative landscape states?: An empirical study on intertidal flats and marshes. *J. Geophys. Res. Earth Surf.* **118**, 229–240 (2013). doi: 10.1029/2012JF002474
12. H. Rydin, J. K. Jeglum, *The Biology of Peatlands* (Oxford Univ. Press, 2013).
13. P. S. Maxwell *et al*., The fundamental role of ecological feedback mechanisms for the adaptive management of seagrass ecosystems—A review. *Biol. Rev. Camb. Philos. Soc.* **92**, 1521–1538 (2017). doi: 10.1111/brv.12294; pmid: 27581168
14. E. B. Barbier *et al*., The value of estuarine and coastal ecosystem services. *Ecol. Monogr.* **81**, 169–193 (2011). doi: 10.1890/10-1510.1
15. J. H. C. Dau, Ansicht einiger der größeren und kleineren Moore Holsteins und Schleswigs, nebst daraus abgeleitete Betrachtungen (Overview of some of the larger and smaller peatlands of Holstein and Schleswig with some considerations derived from them). *Schleswig-Holstein-Lauenburgsche Prov.* **10**, 73–82 (1821).
16. Materials and methods are available as supplementary materials.
17. S. Emerson, J. I. Hedges, Processes controlling the organic carbon content of open ocean sediments. *Paleoceanography* **3**, 621–634 (1988). doi: 10.1029/PA003i005p00621
18. K. MacDicken *et al*., *Global Forest Resources Assessment 2015: How Are the World's Forests Changing?* (Food and Agriculture Organization, ed. 2, 2016).
19. X. Ouyang, S. Y. Lee, Updated estimates of carbon accumulation rates in coastal marsh sediments. *Biogeosciences* **11**, 5057–5071 (2014). doi: 10.5194/bg-11-5057-2014
20. J. J. Middelburg, J. Nieuwenhuize, R. K. Lubberts, O. van de Plassche, Organic carbon isotope systematics of coastal marshes. *Estuar. Coast. Shelf Sci.* **45**, 681–687 (1997). doi: 10.1006/ecss.1997.0247
21. H. Kennedy *et al*., Seagrass sediments as a global carbon sink: Isotopic constraints. *Global Biogeochem. Cycles* **24**, n/a (2010). doi: 10.1029/2010GB003848
22. P. Mueller *et al*., Assessing the long-term carbon-sequestration potential of the semi-natural salt marshes in the European Wadden Sea. *Ecosphere* **10**, e02556 (2019). doi: 10.1002/ecs2.2556
23. S. D. Sasmito *et al*., Organic carbon burial and sources in soils of coastal mudflat and mangrove ecosystems. *Catena* **187**, 104414 (2020). doi: 10.1016/j.catena.2019.104414
24. H. Joosten, D. Clarke, *Wise Use of Mires and Peatlands—Background and Principles Including a Framework for Decision-making* (International Mire Conservation Group and International Peat Society, 2002).
25. L. P. M. Lamers *et al*., Ecological restoration of rich fens in Europe and North America: From trial and error to an evidence-based approach. *Biol. Rev. Camb. Philos. Soc.* **90**, 182–203 (2015). doi: 10.1111/brv.12102; pmid: 24698312
26. J. Schoelynck *et al*., The trapping of organic matter within plant patches in the channels of the Okavango Delta: A matter of quality. *Aquat. Sci.* **79**, 661–674 (2017). doi: 10.1007/s00027-017-0527-2
27. S. F. Stofberg, J. van Engelen, J.-P. M. Witte, S. E. van der Zee, Effects of root mat buoyancy and heterogeneity on floating fen hydrology. *Ecohydrology* **9**, 1222–1234 (2016). doi: 10.1002/eco.1720
28. C. Fritz, D. I. Campbell, L. A. Schipper, Oscillating peat surface levels in a restiad peatland, New Zealand—Magnitude and spatiotemporal variability. *Hydrol. Processes* **22**, 3264–3274 (2008). doi: 10.1002/hyp.6912
29. D. J. Greenwood, The effect of oxygen concentration on the decomposition of organic materials in soil. *Plant Soil* **14**, 360–376 (1961). doi: 10.1007/BF01666294
30. L. P. M. Lamers *et al*., Microbial transformations of nitrogen, sulfur, and iron dictate vegetation composition in wetlands: A review. *Front. Microbiol.* **3**, 156 (2012). doi: 10.3389/fmicb.2012.00156; pmid: 22539932
31. P. Rovira, V. R. Vallejo, Labile and recalcitrant pools of carbon and nitrogen in organic matter decomposing at different depths in soil: An acid hydrolysis approach. *Geoderma* **107**, 109–141 (2002). doi: 10.1016/S0016-7061(01)00143-4
32. N. van Breemen, How Sphagnum bogs down other plants. *Trends Ecol. Evol.* **10**, 270–275 (1995). doi: 10.1016/0169-5347(95)90007-1; pmid: 21237035
33. R. S. Clymo, The origin of acidity in Sphagnum bogs. *Bryologist* **67**, 427–431 (1964). doi: 10.1639/0007-2745(1964)67[427:TOOAIS]2.0.CO;2
34. E. Gorham, S. J. Eisenreich, J. Ford, M. V. Santelmann, in *Chemical Processes in Lakes* (John Wiley and Sons, 1985), pp. 339–362.
35. Z. Yu, J. Loisel, D. P. Brosseau, D. W. Beilman, S. J. Hunt, Global peatland dynamics since the Last Glacial Maximum. *Geophys. Res. Lett.* **37**, 43584 (2010). doi: 10.1029/2010GL043584
36. D. M. Young *et al*., Misinterpreting carbon accumulation rates in records from near-surface peat. *Sci. Rep.* **9**, 17939 (2019). doi: 10.1038/s41598-019-53879-8; pmid: 31784556
37. A. Prager, A. Barthelmes, H. Joosten, A touch of tropics in temperate mires: On Alder carrs and carbon cycles. *Peatlands Int.* **2**, 26–29 (2006).
38. I. Aselmann, P. J. Crutzen, Global distribution of natural freshwater wetlands and rice paddies, their net primary productivity, seasonality and possible methane emissions. *J. Atmos. Chem.* **8**, 307–358 (1989). doi: 10.1007/BF00052709
39. C. M. Duarte, J. Cebrián, The fate of marine autotrophic production. *Limnol. Oceanogr.* **41**, 1758–1766 (1996). doi: 10.4319/lo.1996.41.8.1758
40. S. T. Chew, J. B. Gallagher, Accounting for black carbon lowers estimates of blue carbon storage services. *Sci. Rep.* **8**, 2553 (2018). doi: 10.1038/s41598-018-20644-2; pmid: 29416101
41. D. C. Donato *et al*., Mangroves among the most carbon-rich forests in the tropics. *Nat. Geosci.* **4**, 293–297 (2011). doi: 10.1038/ngeo1123
42. J. L. Breithaupt, J. M. Smoak, T. J. SmithIII, C. J. Sanders, A. Hoare, Organic carbon burial rates in mangrove sediments: Strengthening the global budget. *Global Biogeochem. Cycles* **26**, 2012GB004375 (2012). doi: 10.1029/2012GB004375
43. S. Bouillon *et al*., Mangrove production and carbon sinks: A revision of global budget estimates. *Global Biogeochem. Cycles* **22**, n/a (2008). doi: 10.1029/2007GB003052
44. S. Bouillon, R. M. Connolly, in *Ecological Connectivity Among Tropical Coastal Ecosystems* (Springer, 2009), pp. 45–70.
45. T. J. Bouma *et al*., Trade-offs related to ecosystem engineering: A case study on stiffness of emerging macrophytes. *Ecology* **86**, 2187–2199 (2005). doi: 10.1890/04-1588
46. R. J. M. Temmink *et al*., Mimicry of emergent traits amplifies coastal restoration success. *Nat. Commun.* **11**, 3668 (2020). doi: 10.1038/s41467-020-17438-4; pmid: 32699271
47. M. Van de Broek *et al*., Long-term organic carbon sequestration in tidal marsh sediments is dominated by old-aged allochthonous inputs in a macrotidal estuary. *Glob. Change Biol.* **24**, 2498–2512 (2018). doi: 10.1111/gcb.14089; pmid: 29431887
48. K. Koop-Jakobsen, F. Wenzhöfer, The dynamics of plant-mediated sediment oxygenation in *Spartina anglica* rhizospheres—A planar optode study. *Estuaries Coasts* **38**, 951–963 (2015). doi: 10.1007/s12237-014-9861-y
49. M. E. Gonneea, A. Paytan, J. A. Herrera-Silveira, Tracing organic matter sources and carbon burial in mangrove sediments over the past 160 years. *Estuar. Coast. Shelf Sci.* **61**, 211–227 (2004). doi: 10.1016/j.ecss.2004.04.015
50. R. K. James *et al*., Maintaining tropical beaches with seagrass and algae: A promising alternative to engineering solutions. *Bioscience* **69**, 136–142 (2019). doi: 10.1093/biosci/biy154
51. G. S. Fivash *et al*., Elevated micro-topography boosts growth rates in Salicornia procumbens by amplifying a tidally-driven oxygen pump: Implications for natural recruitment and restoration. *Ann. Bot.* **125**, 353–364 (2019). doi: 10.1093/aob/mcz137
52. T. B. Atwood *et al*., Predators help protect carbon stocks in blue carbon ecosystems. *Nat. Clim. Chang.* **5**, 1038–1045 (2015). doi: 10.1038/nclimate2763
53. L. Ren, K. Jensen, P. Porada, P. Mueller, Biota-mediated carbon cycling-A synthesis of biotic-interaction controls on blue carbon. *Ecol. Lett.* **25**, 521–540 (2022). doi: 10.1111/ele.13940; pmid: 35006633
54. L. G. Gillis *et al*., Potential for landscape-scale positive interactions among tropical marine ecosystems. *Mar. Ecol. Prog. Ser.* **503**, 289–303 (2014). doi: 10.3354/meps10716
55. C. J. T. Ladd, M. F. Duggan-Edwards, T. J. Bouma, J. F. Pages, M. W. Skov, Sediment supply explains long-term and large-scale patterns in salt marsh lateral expansion and erosion. *Geophys. Res. Lett.* **46**, 11178–11187 (2019). doi: 10.1029/2019GL083315
56. C. M. Duarte, I. J. Losada, I. E. Hendriks, I. Mazarrasa, N. Marbà, The role of coastal plant communities for climate change mitigation and adaptation. *Nat. Clim. Chang.* **3**, 961–968 (2013). doi: 10.1038/nclimate1970
57. M. L. Kirwan, J. P. Megonigal, Tidal wetland stability in the face of human impacts and sea-level rise. *Nature* **504**, 53–60 (2013). doi: 10.1038/nature12856; pmid: 24305148
58. M. L. Kirwan, S. Temmerman, E. E. Skeehan, G. R. Guntenspergen, S. Fagherazzi, Overestimation of marsh vulnerability to sea level rise. *Nat. Clim. Chang.* **6**, 253–260 (2016). doi: 10.1038/nclimate2909
59. A. Jacotot, C. Marchand, B. E. Rosenheim, E. W. Domack, M. Allenbach, Mangrove sediment carbon stocks along an elevation gradient: Influence of the late Holocene marine regression (New Caledonia). *Mar. Geol.* **404**, 60–70 (2018). doi: 10.1016/j.margeo.2018.07.005
60. J. Leifeld, L. Menichetti, The underappreciated potential of peatlands in global climate change mitigation strategies. *Nat. Commun.* **9**, 1071 (2018). doi: 10.1038/s41467-018-03406-6; pmid: 29540695
61. K. B. Gedan, B. R. Silliman, M. D. Bertness, Centuries of human-driven change in salt marsh ecosystems. *Ann. Rev. Mar. Sci.* **1**, 117–141 (2009). doi: 10.1146/annurev.marine.010908.163930; pmid: 21141032
62. M. Waycott *et al*., Accelerating loss of seagrasses across the globe threatens coastal ecosystems. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 12377–12381 (2009). doi: 10.1073/pnas.0905620106; pmid: 19587236
63. Millennium Ecosystem Assessment, *Ecosystems and Human Well-Being: Synthesis* (Island Press, 2005).
64. R. J. Orth *et al*., A Global Crisis for Seagrass Ecosystems. *Bioscience* **56**, 987–996 (2006). doi: 10.1641/0006-3568(2006)56[987:AGCFSE]2.0.CO;2
65. M. F. Adame *et al*., Future carbon emissions from global mangrove forest loss. *bioRxiv*15571 [Preprint] (2020). doi: 10.1101/2020.08.27.271189
66. C. E. Lovelock, R. Reef, Variable impacts of climate change on blue carbon. *One Earth* **3**, 195–211 (2020). doi: 10.1016/j.oneear.2020.07.010
67. M. R. Turetsky *et al*., Global vulnerability of peatlands to fire and carbon loss. *Nat. Geosci.* **8**, 11–14 (2015). doi: 10.1038/ngeo2325
68. G. Hugelius *et al*., Large stocks of peatland carbon and nitrogen are vulnerable to permafrost thaw. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 20438–20446 (2020). doi: 10.1073/pnas.1916387117; pmid: 32778585
69. M. R. Turetsky *et al*., Carbon release through abrupt permafrost thaw. *Nat. Geosci.* **13**, 138–143 (2020). doi: 10.1038/s41561-019-0526-0
70. K. Hergoualc'h, L. V. Verchot, Stocks and fluxes of carbon associated with land use change in Southeast Asian tropical peatlands: A review. *Global Biogeochem. Cycles* 10.1029/2009GB003718 (2011). doi: 10.1029/2009GB003718
71. B. Tiemeyer *et al*., High emissions of greenhouse gases from grasslands on peat and other organic soils. *Glob. Change Biol.* **22**, 4134–4149 (2016). doi: 10.1111/gcb.13303; pmid: 27029402
72. A. Günther *et al*., Prompt rewetting of drained peatlands reduces climate warming despite methane emissions. *Nat. Commun.* **11**, 1644 (2020). doi: 10.1038/s41467-020-15499-z; pmid: 32242055
73. Z. Wang *et al*., Human-induced erosion has offset one-third of carbon emissions from land cover change. *Nat. Clim. Chang.* **7**, 345–349 (2017). doi: 10.1038/nclimate3263
74. L. Pendleton *et al*., Estimating global "blue carbon" emissions from conversion and degradation of vegetated coastal ecosystems. *PLOS ONE* **7**, e43542 (2012). doi: 10.1371/journal.pone.0043542; pmid: 22962585
75. P. I. Macreadie *et al*., Blue carbon as a natural climate solution. *Nat. Rev. Earth Environ.* **2**, 826–839 (2021). doi: 10.1038/s43017-021-00224-1
76. E. Bayraktarov *et al*., The cost and feasibility of marine coastal restoration. *Ecol. Appl.* **26**, 1055–1074 (2016). doi: 10.1890/15-1077; pmid: 27509748
77. R. S. DE Groot *et al*., Benefits of investing in ecosystem restoration. *Conserv. Biol.* **27**, 1286–1293 (2013). doi: 10.1111/cobi.12158; pmid: 24112105
78. E. Romijn *et al*., Land restoration in Latin America and the Caribbean: An overview of recent, ongoing and planned restoration initiatives and their potential for climate change mitigation. *Forests* **10**, 510 (2019). doi: 10.3390/f10060510
79. R. Andersen *et al*., An overview of the progress and challenges of peatland restoration in Western Europe. *Restor. Ecol.* **25**, 271–282 (2017). doi: 10.1111/rec.12415

80. T. J. Bouma *et al*., Density-dependent linkage of scale-dependent feedbacks: A flume study on the intertidal macrophyte Spartina anglica. *Oikos* **118**, 260–268 (2009). doi: 10.1111/j.1600-0706.2008.16892.x
81. M. Scheffer, S. Carpenter, J. A. Foley, C. Folke, B. Walker, Catastrophic shifts in ecosystems. *Nature* **413**, 591–596 (2001). doi: 10.1038/35098000; pmid: 11595939
82. T. J. Bouma *et al*., Identifying knowledge gaps hampering application of intertidal habitats in coastal protection: Opportunities and steps to take. *Coast. Eng.* **87**, 147–157 (2014). doi: 10.1016/j.coastaleng.2013.11.014
83. B. R. Silliman *et al*., Facilitation shifts paradigms and can amplify coastal restoration efforts. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 14295–14300 (2015). doi: 10.1073/pnas.1515297112; pmid: 26578775
84. M. M. Katwijk *et al*., Global analysis of seagrass restoration: The importance of large-scale planting. *J. Appl. Ecol.* **53**, 567–578 (2016). doi: 10.1111/1365-2664.12562
85. T. van der Heide *et al*., Coastal restoration success via emergent trait-mimicry is context dependent. *Biol. Conserv.* **264**, 109373 (2021). doi: 10.1016/j.biocon.2021.109373
86. G. S. Fivash *et al*., Restoration of biogeomorphic systems by creating windows of opportunity to support natural establishment processes. *Ecol. Appl.* **31**, e02333 (2021). doi: 10.1002/eap.2333; pmid: 33768651
87. N. Kettridge *et al*., Moderate drop in water table increases peatland vulnerability to post-fire regime shift. *Sci. Rep.* **5**, 8063 (2015). doi: 10.1038/srep08063; pmid: 25623290
88. W. Wichtmann, C. Schröder, H. Joosten, *Paludiculture-Productive Use of Wet Peatlands* (Schweizerbart Science Publishers, 2016).
89. C. D. Evans *et al*., Overriding water table control on managed peatland greenhouse gas emissions. *Nature* **593**, 548–552 (2021). doi: 10.1038/s41586-021-03523-1; pmid: 33882562
90. H. Joosten, *The Global Peatland CO_2 Picture: Peatland Status and Drainage Related Emissions in all Countries of the World* (Univ. Greifswald, 2009).
91. H. Joosten, M. Tapio-Biström, S. Tol, *Peatlands—Guidance for Climate Change Mitigation Through Conservation, Rehabilitation and Sustainable Use* (Food and Agriculture Organization of the United Nations and Wetlands International, 2012).
92. J. Sanderman *et al*., A global map of mangrove forest soil carbon at 30 m spatial resolution. *Environ. Res. Lett.* **13**, 055002 (2018). doi: 10.1088/1748-9326/aabe1c
93. C. J. Mcowen *et al*., A global map of saltmarshes. *Biodivers. Data J.* **5**, e11764 (2017). doi: 10.3897/BDJ.5.e11764; pmid: 28765720
94. E. P. Green, F. T. Short, T. Frederick, *World Atlas of Seagrasses* (Univ. California Press, 2003).
95. R. J. M. Temmink *et al*., Data from: Recovering wetland biogeomorphic feedbacks to restore the world's biotic carbon hotspots. *Zenodo* (2022); doi: 10.5281/zenodo.6411400

ACKNOWLEDGMENTS

The authors thank D. de Groot for his valuable comments on the manuscript and T. A. W. Markus for designing the figures. Symbols for diagrams are courtesy of the Integration and Application Network, IAN Image Library (https://ian.umces.edu/imagelibrary). **Funding:** R.J.M.T. was funded by NWO/TTW-OTP grant 14424; T.v.d.H. was funded by NWO/TTW-Vidi grant 16588; and B.R.S. was funded by grants from the Oak Foundation, Duke RESTORE, and the Lenfest Ocean Program. C.A. was supported by NSF CAREER (#1652628). C.F. was funded by EC H2020 grant 101056848 and NWO ALW.GAS.4. T.J.B. thanks Natuurmonumenten for financial support. **Author contributions:** All authors contributed to the conceptualization and research ideas. R.J.M.T., L.P.M.L., and T.v.d.H. wrote the original draft, and all authors contributed to reviewing and editing the paper. R.J.M.T. and T.v.d.H. prepared visual materials. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Data underlying analyses presented in this study are deposited in Zenodo at (*95*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abn1479
Materials and Methods
Tables S1 and S2
References (*96*–*139*)

Submitted 6 November 2021; accepted 5 April 2022
10.1126/science.abn1479



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

NEUROSCIENCE

The geometry of domain-general performance monitoring in the human medial frontal cortex

Zhongzheng Fu*, Danielle Beam, Jeffrey M. Chung, Chrystal M. Reed, Adam N. Mamelak, Ralph Adolphs, Ueli Rutishauser*

INTRODUCTION: Monitoring our own behavior without explicit feedback is a prominent human ability, enabling us to evaluate whether we made an error, are experiencing conflict, or are in an easy or difficult task environment. In real-life situations, we often need to rapidly learn to perform novel tasks with minimal instruction and time for practice, which requires domain-general performance monitoring that functions "out of the box." At the same time, we need to resolve unforeseen errors, difficulty, and other performance disturbances with task-specific measures ad hoc. Doing so requires domain-specific performance monitoring processes so as to improve upon and specialize task performance. Are there neural representations that support both domain-specific and domain-general performance monitoring, and if so, how are these requirements satisfied simultaneously?

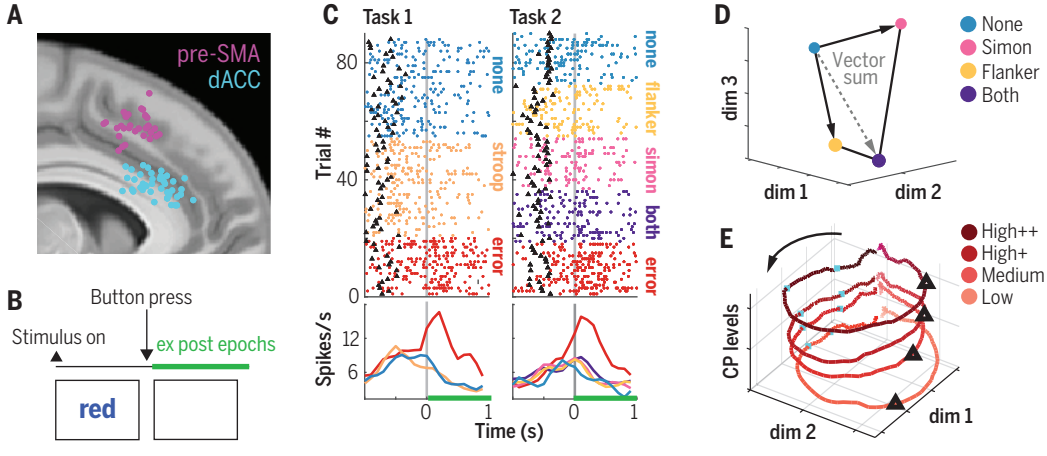
RATIONALE: In this work, we focused on the medial frontal cortex (MFC), which is known to serve a central role in performance monitoring. We recorded from single neurons in the MFC while neurosurgical patients performed two tasks in which errors were caused by different kinds of conflict. We identified neurons whose responses encoded performance monitoring variables in one or both tasks and characterized the resulting high-dimensional neural representations supported by neuronal populations. Although specialization and generalization place different constraints on the structure of neural representations, our work suggests that the geometry of population activity can be configured to accommodate both seemingly conflicting demands.

RESULTS: We recorded from 1431 single neurons in the MFC [dorsal anterior cingulate cortex (dACC) and pre–supplementary motor area (pre-SMA)] across 34 patients. Subjects were asked to press, as quickly as possible, one of three buttons to indicate the ink color of the word "red," "green," or "blue" (task 1, Stroop task) or to report the unique number among a string of numbers (task 2, multisource interference task). We modeled serial performance of both tasks using a hierarchical Bayesian framework. Our models assumed that subjects maintained an internal estimate of the probability of encountering a certain type of conflict (conflict probability, or CP) and iteratively updated this estimate using Bayes' law. Reaction times were generated by a drift-diffusion model, with CP biasing the drift rates. These behavioral analyses showed that the reaction times and likelihood of making an error varied systematically with trial history. At the single-neuron level, cells encoded different types of conflict, CP, and/or errors in one or both tasks. Individual neurons thus demonstrated heterogeneous task specificity when encoding these performance-monitoring variables, precluding a simplistic interpretation of domain generality at the single-neuron level. At the population level, however, these cells formed a high-dimensional representation with a geometry that allowed task-invariant decoding of all three variables on single trials, while at the same time allowing decoding of task-specific performance monitoring variables. Representations of conflict probability were consistent with that expected of the dynamics of a line attractor, with levels of CP stably maintained throughout the trial. The neural states indicating the presence of multiple types of conflict were equal to the vector sum of the states indicating each individual type of conflict, thereby revealing a compositional representation. Lastly, retrospective representations of conflict served to update internal estimates of conflict probability.

CONCLUSION: We leveraged the opportunity to record from the same populations of single neurons in the human MFC during two tasks to identify the structure of neural representations supporting performance monitoring. Population activity could be factorized into task identity and task-invariant dimensions that were orthogonal to each other. This geometry of the population activity could allow downstream brain regions to read out both domain-general and domain-specific signals from the same group of neurons and to initiate corresponding physiological and/or behavioral adaptations. These findings reveal how representations of evaluative signals can be both abstract and task-specific and suggest a neuronal mechanism for estimating control demand. ■

The list of author affiliations is available in the full article online.
*Corresponding author. Email: rutishauseru@csmc.edu (U.R.); zzbrooksfu@gmail.com (Z.F.)
Cite this article as Z. Fu *et al.*, *Science* **376**, eabm9922 (2022). DOI: 10.1126/science.abm9922

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abm9922



Representations of evaluative signals in the human frontal cortex are both abstract and task-specific. (**A**) Recording locations. (**B**) Analysis epochs. (**C**) Response of example neuron in both tasks, demonstrating domain generality for errors (red) and no differentiation of different types of correct conflict trials (all other colors). (**D**) Compositionality of population-level conflict representations. (**E**) Conflict probability displaces dynamics in neural state space.

RESEARCH ARTICLE

NEUROSCIENCE

The geometry of domain-general performance monitoring in the human medial frontal cortex

Zhongzheng Fu[1,2]*, Danielle Beam[1]†, Jeffrey M. Chung[3], Chrystal M. Reed[3], Adam N. Mamelak[1], Ralph Adolphs[2,4], Ueli Rutishauser[1,3,4,5]*

Controlling behavior to flexibly achieve desired goals depends on the ability to monitor one's own performance. It is unknown how performance monitoring can be both flexible, to support different tasks, and specialized, to perform each task well. We recorded single neurons in the human medial frontal cortex while subjects performed two tasks that involve three types of cognitive conflict. Neurons encoding conflict probability, conflict, and error in one or both tasks were intermixed, forming a representational geometry that simultaneously allowed task specialization and generalization. Neurons encoding conflict retrospectively served to update internal estimates of conflict probability. Population representations of conflict were compositional. These findings reveal how representations of evaluative signals can be both abstract and task-specific and suggest a neuronal mechanism for estimating control demand.

Humans can rapidly learn to perform novel tasks given abstract rules, even if task requirements differ drastically. To achieve this, cognitive control must coordinate processes across a diverse array of perceptual, motor, and memory domains and at different levels of abstraction over sensorimotor representations (*1*–*3*). A key component of cognitive control is performance monitoring, which enables us to evaluate whether we have made an error, experienced conflict, and responded quickly or slowly (*4*, *5*). It provides task-specific information about which processes cause an error or a slow response so that they can be selectively guided (*6*–*16*). At the same time, performance monitoring needs to be flexible and domain-general to enable cognitive control for novel tasks (*17*), to inform abstract strategies [e.g., "win–stay, lose–switch," exploration versus exploitation (*18*, *19*)], and to initiate global adaptations (*20*–*28*). As an example, errors and conflicts can have different causes in different tasks (task-specific) but all signify failure or difficulty to fulfill an intended abstract goal (task-general); performance monitoring should satisfy both requirements. Thanks to its broad connectivity (*29*, *30*), the medial frontal cortex (MFC) serves a central role in evaluating one's own performance and decisions (*4*, *10*, *13*, *16*, *19*, *24*, *31*–*43*). However, little is known about how neural representation in the MFC can support both domain-specific and domain-general adaptations.

Specialization and generalization place different constraints on neural representations (*44*, *45*). Specialization demands separation of encoded task parameters, which can be fulfilled by increasing the dimensionality of neural representations (*46*, *47*). By contrast, generalization involves abstracting away details specific to performing a single task, which can be achieved by reducing the representational dimensionality (*3*, *46*). Theoretical work shows that the geometry of population activity can be configured to accommodate both of these seemingly conflicting demands (*44*), provided that the constituent single neurons multiplex task parameters nonlinearly (*48*, *49*). Although recent experimental work has shown that neuronal population activity is indeed organized this way in the frontal cortex and hippocampus in macaques (*44*) and humans (*50*) during decision-making tasks, it remains unknown whether this framework is applicable to the important topic of cognitive control.

A key aspect of behavioral control is learning about the identity and intensity of control needed to correctly perform a task and deploy control proactively on the basis of such estimates (*13*, *51*). This requires integrating internally generated performance outcomes over multiple trials. Blood oxygen level–dependent functional magnetic resonance imaging (BOLD-fMRI) studies localize signals related to control demand estimation to the insular-frontostriatal network (*52*). However, the neuronal mechanisms of how these estimates are updated trial by trial and whether the underlying substrate is domain-general or task-specific remain unknown.

Results

Task and behavior

Subjects (see table S1) performed the multi-source interference task (MSIT) and the color-word Stroop task (Fig. 1A and methods in the supplementary materials). Conflict and errors arose from different sources in these two tasks: competition between prepotency of reading and color naming in the Stroop task, and competition between the target response and either the spatial location of target ("Simon effect," denoted by "si") or flanking numbers ("flanker effect," denoted by "fl"), or both ("sf") in the MSIT. In the MSIT, we refer to trials with or without a Simon conflict as "Simon" and "non-Simon" trials, respectively (and similarly for flanker trials). Stimulus sequences were randomized, and each type of trial occurred with a fixed probability (see methods for details).

Subjects performed well (Stroop error rate: 6.3 ± 5%; MSIT error rate: 6.0 ± 5%). Reaction times (RTs) on correct trials were significantly prolonged in the presence of conflicts (Fig. 1B; fig. S1, D and E, shows raw RTs). Participants' successive performance (RT and accuracy) was modeled with a hierarchical Bayesian model (*52*–*54*). The model (Fig. 1C) assumes that participants iteratively updated internal estimates of how likely they were to encounter a certain type of conflict trial ("conflict probability") on the basis of their previous estimates and new evidence (experienced conflict and RT) on the current trial using Bayes' law. Given that trial sequences were randomized by the experimenter, subjects could not predict with certainty whether an upcoming trial would have conflict or not, but could instead estimate the probability of a conflict trial, which is fixed a priori but unbeknownst to the subject. The decision process was modeled as a drift-diffusion process (DDM), with drift rates being a function of conflict and subjects' estimated conflict probability ("CP coeff" in Fig. 1E). Model parameters (DDM parameters and trial-wise CP estimation) were inferred from subjects' behavioral data (RT and trial outcomes) and conflict sequences using data from all sessions and full Bayesian inference (see methods for details). Significance of model parameters was determined from the posterior distributions directly (see methods). The posterior predictive distributions captured the true RT distribution well (fig. S2, A and B). The Bayesian models were identifiable (fig. S2, C and D). Figure 1D shows RT and estimated conflict probability of an example MSIT session. Unless specified otherwise, we use the term "conflict probability" (or CP) to refer to

[1]Department of Neurosurgery, Cedars-Sinai Medical Center, Los Angeles, CA, USA. [2]Division of Humanities and Social Sciences, California Institute of Technology, Pasadena, CA, USA. [3]Department of Neurology, Cedars-Sinai Medical Center, Los Angeles, CA, USA. [4]Division of Biology and Bioengineering, California Institute of Technology, Pasadena, CA, USA. [5]Center for Neural Science and Medicine, Department of Biomedical Sciences, Cedars-Sinai Medical Center, Los Angeles, CA, USA.
***Corresponding author. Email: rutishauseru@csmc.edu (U.R.); zzbrooksfu@gmail.com (Z.F.)**
†Present address: Grossman School of Medicine, New York University, New York, NY, USA.

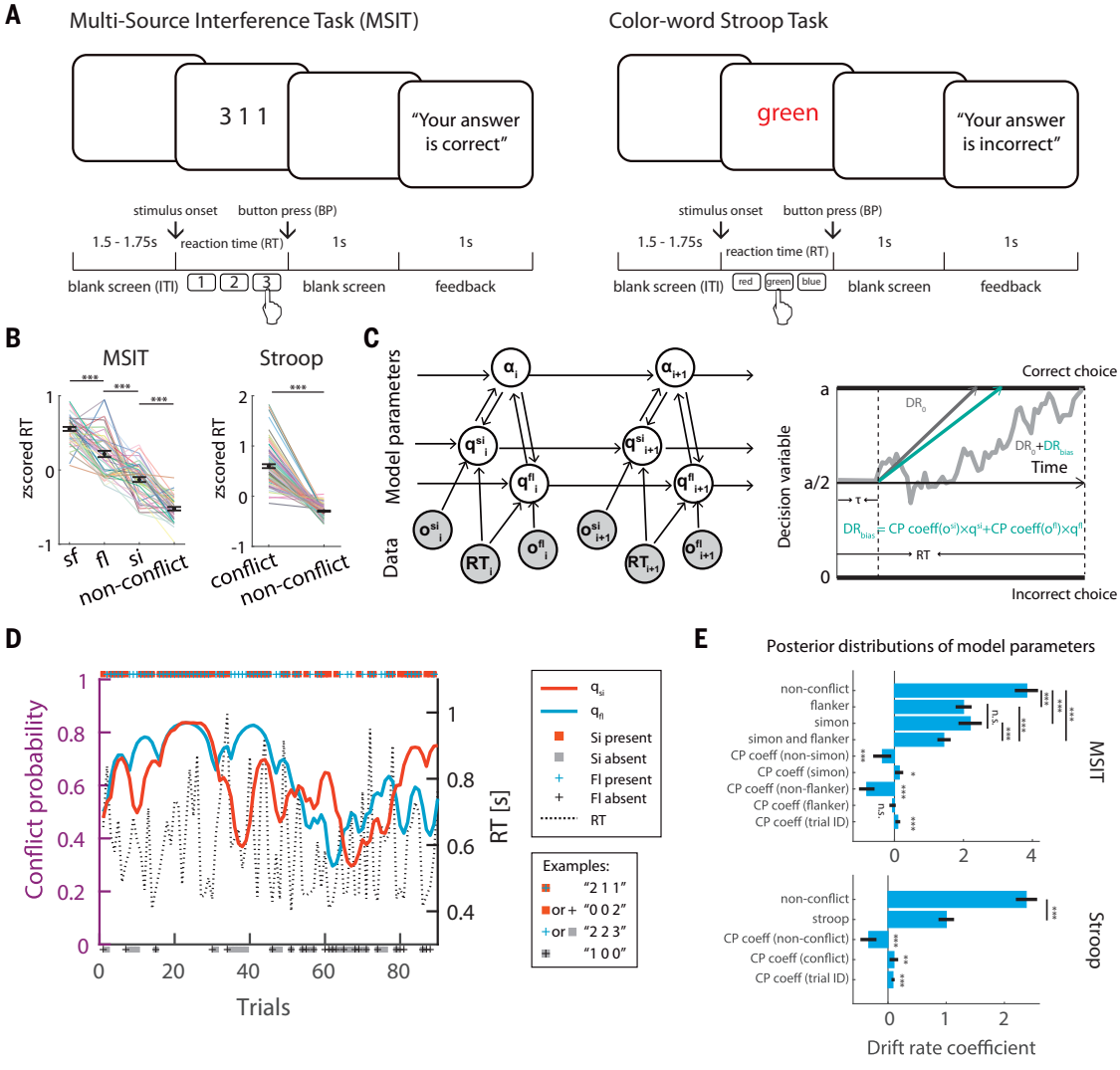


Fig. 1. Tasks, Bayesian conflict learning model, and reaction time analyses. (**A**) Task structure. (**B**) RTs were significantly prolonged by conflict in MSIT (left, n = 41 sessions) and Stroop task (right, n = 82 sessions). (**C**) The conflict probability estimation process (left) and the decision process modeled as a drift diffusion (right). Shown is the MSIT model, which has the six variables learning rate (α), Simon probability (q_{si}), flanker probability (q_{fl}), observed Simon conflict (o_{si}), observed flanker conflict (o_{fl}), and RT. Observables (trial congruency, RT, and outcome) are shown in gray; model parameters are in white. Arrows indicate information flow. (**D**) Estimated Simon probability (red) and flanker probability (blue) from an example MSIT session. Markers along the top indicate the type of conflict present. (**E**) Posterior distributions of model parameters after fitting to the behavior of all subjects. Black bars show high-density intervals. CP had a significant effect on RT. Vertical bars with asterisks show comparisons between posterior distribution, and asterisks without vertical bars mark comparisons with zero. *P < 0.05; **P < 0.01; ***P < 0.001; n.s., not significant (P > 0.05).

the posterior means of the estimated conflict probability on each trial.

Conflicts reduced drift rates and thus prolonged RT [$P(dr_{\text{Stroop nonconflict}} < dr_{\text{Stroop conflict}}) < 0.001$, $P(dr_{\text{MSIT nonconflict}} < dr_{\text{Simon}}) < 0.001$, $P(dr_{\text{MSIT nonconflict}} < dr_{\text{flanker}}) < 0.001$, $P(dr_{\text{MSIT nonconflict}} < dr_{\text{Simon and flanker}}) < 0.001$]. High estimated CP prolonged RT on non-conflict trials [negative CP coefficient decreases drift rate: $P(CP\ coeff_{\text{Stroop nonconflict}} > 0) < 0.001$, $P(CP\ coeff_{\text{non-Simon}} > 0) < 0.001$, $P(CP\ coeff_{\text{nonflanker}} > 0) < 0.001$] but hastened RT on conflict trials [positive CP coefficient increases drift rate: $P(CP\ coeff_{\text{Stroop}} < 0) < 0.01$, $P(CP\ coeff_{\text{Simon}} < 0) < 0.05$] (Fig. 1E). Our model yielded qualitatively similar results when applied to data from a separate group of healthy control subjects (fig. S1C). In a separate control model where only trial outcomes and conflict sequences were used (but not RT), high CP correlated with reduced likelihood of making an error on conflict trials [Bayesian logistic regression; negative CP coefficient reduces error likelihood: $P(CP\ coeff_{\text{Stroop}} > 0) < 0.01$, $P(CP\ coeff_{\text{Simon}} > 0) < 0.05$] (fig. S1B), suggesting that higher CP predicted higher levels of control. See (*55*) for more about model comparisons.

Neuronal correlates of performance-monitoring signals

We collected single-neuron recordings from two regions within the MFC (Fig. 2A): the dorsal anterior cingulate cortex (dACC) and the pre–supplementary motor area (pre-SMA). We isolated 1431 putative single neurons [Stroop: 584 in the dACC and 607 in the pre-SMA across 34 participants (10 females); MSIT: 326 in the dACC and 412 in the pre-SMA in 12 participants (6 females)]. Neurons in the dACC and the pre-SMA were pooled, unless specified otherwise, because neurons responded similarly (but see the section "Comparison between the dACC and the pre-SMA" below for notable differences).

We identified neurons selective for the mean and variance of the posterior distribution of CP, error, and conflict surprise (unsigned prediction error of CP: 1 – |*conflict* – *CP*|, where *conflict* is an indicator variable) in the ex post epoch, and conflict in the ex ante and ex post epochs (see example neurons in Fig. 2, schematic of analysis epochs in Fig. 3A, and a summary of overall cell counts in Fig. 3B).

A substantial proportion of neurons encoded the mean and variance of CP posterior distribution in the ex post epoch (Fig. 3B; 25 and 17% of neurons in MSIT, 21 and 12% in Stroop). CP is maintained up to the baseline of the next trial [Fig. 3B, "conflict prob. (baseline)", blue; 21% in MSIT, 17% in Stroop]. Neurons encoded conflict in the ex ante epoch (17 and 14% of neurons in MSIT and Stroop, respectively;

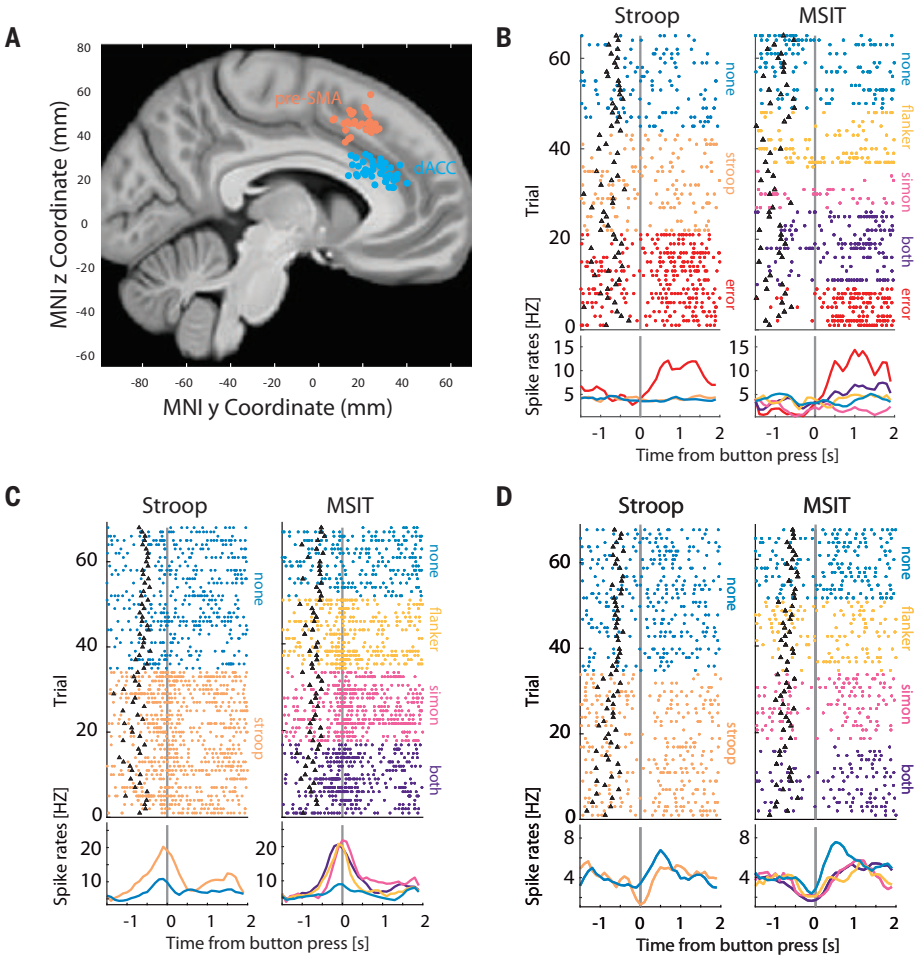


Fig. 2. Recording locations and example neurons. (**A**) Recording locations shown on the CIT168 Atlas brain. Each dot indicates the location of a microwire bundle. MNI, Montreal Neurological Institute. (**B** to **D**) Activity of three example neurons that show similar response dynamics in both tasks: neuron signaling action error (B), conflict by firing rate increase (C), and conflict by firing rate decrease (D). The black triangles mark stimulus onset. Trials are resorted by type and subsampled to equalize trial numbers for visualization only.

Fig. 3B, cyan), consistent with previous reports (*33*, *38*), but, intriguingly, also in the ex post epoch (21% in MSIT, 20% in Stroop). Neurons encoded conflict surprise (16% in MSIT, 12% in Stroop) and errors (20% in MSIT, 30% in Stroop; see Fig. 2B for an example) in the ex post epoch. The distributions of selective units were similar between the two tasks (Fig. 3B).

Approximately 30% of conflict neurons were selective exclusively in either the ex ante, early (0 to 1 s after button presses), or late (1 to 2 s after button presses) ex post epochs (Fig. 3C), with some (18%) selective in the ex ante, early, and late ex post epochs ("extended"). Distributions of conflict signals across time were similar between the MSIT and Stroop tasks (Fig. 3C). We were intrigued by neurons signaling conflict ex post (20% of neurons in both tasks; Fig. 2, C and D, shows examples), which has not been reported before. This conflict signal, which arises too late to be useful for within-trial cognitive control, was more prominent than the one found in the ex ante epoch in both tasks [15% versus 20%; $\chi^2(1) = 18.34$, $P < 0.001$, chi-squared test].

Many conflict neurons also signaled errors, surprise, CP, or a conjunction of these variables (Fig. 3D). This multiplexing depended on the timing of conflict signals. The proportion of conflict neurons that multiplexed CP (light green bars) increased significantly toward the end of the trial, when updating would be mostly complete [proportion in the late ex post epoch versus that in all other epochs; $\chi^2(1) = 6.86$, $P = 0.008$ for MSIT; $\chi^2(1) = 3.6$, $P = 0.04$ for Stroop, chi-squared test]. Consistent with this idea, the group of neurons signaling conflict exclusively in the ex ante epoch showed the least multiplexing, consistent with a primary role in monitoring conflict during action production [proportion of "pure" conflict neurons active only during the ex ante epoch versus those that are active in other epochs; $\chi^2(1) = 4.93$, $P = 0.03$ for MSIT; $\chi^2(1) = 9.36$, $P = 0.002$ for Stroop, chi-squared test].

We next investigated the temporal stability of the multivariate coding patterns for the performance monitoring variables in each task. We trained a decoder with data from all recorded neurons in one epoch and tested it on data from a different epoch. RTs were equalized across conditions for this analysis. Error decoders trained on early ex post data generalized poorly to later epochs, but the ones trained on late ex post data generalized well into earlier epochs (fig. S6, A and B), suggesting that some neurons represented errors only late in the trial, with a possible role in post-error adjustments (*33*). The multivariate coding patterns for conflict changed substantially to prevent generalization between the ex post and ex ante epochs (fig. S6, C to E; green horizontal lines overlapped with orange and blue only minimally), demonstrating that the MFC encoded these two types of conflict information with different populations of neurons. The magnitude of CP (quantized to allow construction of pseudo-populations) could also be decoded trial by trial from population activity. Decoders trained to decode CP before updating (constructed with baseline spike counts) generalized well to decode CP after updating (using ex post spike counts), and vice versa, suggesting that CP representation was stably maintained (fig. S7). Representation of previous trial conflict (indicator; "1" denotes previous trial had conflict) was relatively weak (fig. S8, A to C; <60% decoding accuracy), consistent with the model comparison results where previous conflict alone was a poor predictor of RT compared with estimated CP (tables S2 and S3). CP neurons demonstrated the greatest degree of multiplexing (Fig. 3F). Only 18% of these neurons signaled CP exclusively, with the remainder multiplexing information about estimated CP on the last trial, conflict, conflict surprise, or a conjunction of these variables. CP neurons changed their firing rates ex post by an amount commensurate with the numerical change of conflict probability between contiguous trials (Fig. 3G, $P < 0.001$, t test against zero; mean correlations in Simon, flanker, and Stroop are 0.036, 0.036, and 0.045, respectively), reflecting a neural "updating" process. Although for some neurons the updating process started in the ex ante epoch, for the majority it started in the ex post epoch and continued well into the ex post period (fig. S3, C to E, shows time courses).

CP neurons differed qualitatively from those of other recorded neurons. First, their trial-by-trial baseline spike rates (treated as a time series) had higher self-similarity than all other types of neurons studied [fig. S4, A and B; self-similarity quantified by α values from the detrended fluctuation analysis (DFA); fig. S4, C to F,

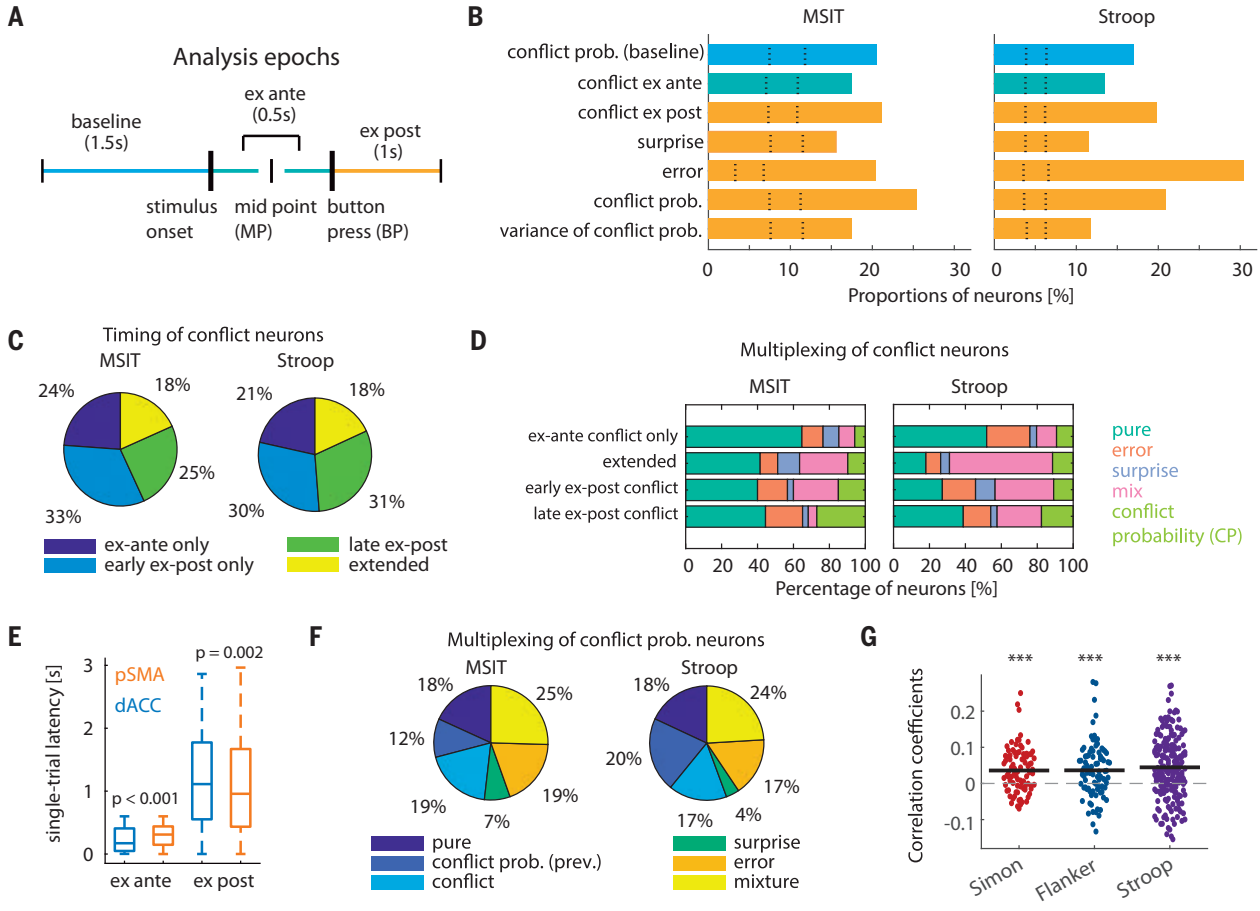


Fig. 3. Single-neuron tuning properties. (**A**) Illustration of epochs used for analysis. Thick vertical bars represent physical events, the slim vertical bars demarcate epochs. (**B**) Percentage of neurons encoding the variable indicated in the two tasks. The color code is as indicated in (A). Dotted lines represent 2.5th and 97.5th percentiles of the null distribution obtained from permutation. For all groups shown, $P < 0.001$. (**C**) Percentage of conflict neurons that are selective in each time period. Early and late ex post epochs denote 0 to 1 s and 1 to 2 s after button press, respectively. (**D**) Percentage of conflict neurons that were also selective for error, surprise, CP, or any combination of these factors ("mix"). (**E**) Comparison of single-trial neuronal response latency of conflict neurons in dACC and pre-SMA (correct trials only, $t = 0$ is stimulus onset for ex ante and button press for ex post conflict neurons). (**F**) Percentage of conflict probability neurons that were also selective for CP on the last trial, conflict, surprise, error, or all combinations of these variables ("mix"). (**G**) Neuronal signature of updating conflict probability estimation. Correlation is computed between the difference between current estimation and conflict probability on the last trial (behavioral update) and the difference between demeaned $FR_{ex\ post}$ and $FR_{baseline}$ (neural update) for all conflict probability neurons. ***$P < 0.001$.

shows DFA analysis of two example neurons]. Second, the extracellular spikes generated by these neurons had significantly narrower widths compared with other neurons (fig. S4, G and H, right panels; more broadly, DFA α is negatively correlated with spike width, shown in fig. S4, G and H, left panels).

Representational geometry of conflict within a single task

We next investigated the representational geometry of conflict in the high-dimensional neural state space formed by all recorded neurons. This is possible because different types of cognitive conflict coexist within MSIT. We first tested whether the four MSIT conflict conditions (si, fl, sf, and nonconflict) are separable in the neural state space. A "coding dimension" was defined as the high-dimensional vector flanked by the means of the sf and nonconflict ("none") trials (Fig. 4A, broken lines; see methods). Held-out single-trial data projected onto this coding dimension allowed a decoder to differentiate with high accuracy between all pairs of conflict conditions in the ex ante epoch (Fig. 4B, left), and all but one pair (si versus sf, $P = 0.16$) in the ex post epoch (Fig. 4B, right). Notably, si and fl conflict could be differentiated above chance by this coding dimension. Conflict conditions were still separable after equalizing for RT across conditions (fig. S9C; see methods for RT equalization), suggesting that the separation was independent of trial difficulty for which RT is a proxy (*56*). We next investigated whether representations generalized between Simon and flanker conflict. For each time bin, a coding dimension was constructed by connecting, in the neural state space, the mean of Simon (si + sf) trials with the mean of nonSimon (fl + none) trials, and a separate one by connecting the flanker (fl + sf) versus nonflanker (si + none) means. Held-out single-trial data from either conflict type projected onto the coding dimension constructed using data from another conflict type allowed decoding with high accuracy (Fig. 4C; gray trace, flanker coding dimension decoding Simon versus non-Simon; black trace, Simon coding dimension decoding flanker versus nonflanker).

We next tested whether the representation of conflict was compositional. If true, relative to the mean of "none" trials, the sf representation should be located where the linear vector sum of Simon (si) and flanker (fl) trials lands, forming a parallelogram with the none-sf axis at its diagonal (Fig. 4A). To test this with a decoding approach, a decoder trained to differentiate between the two classes connected by one edge of this parallelogram should be

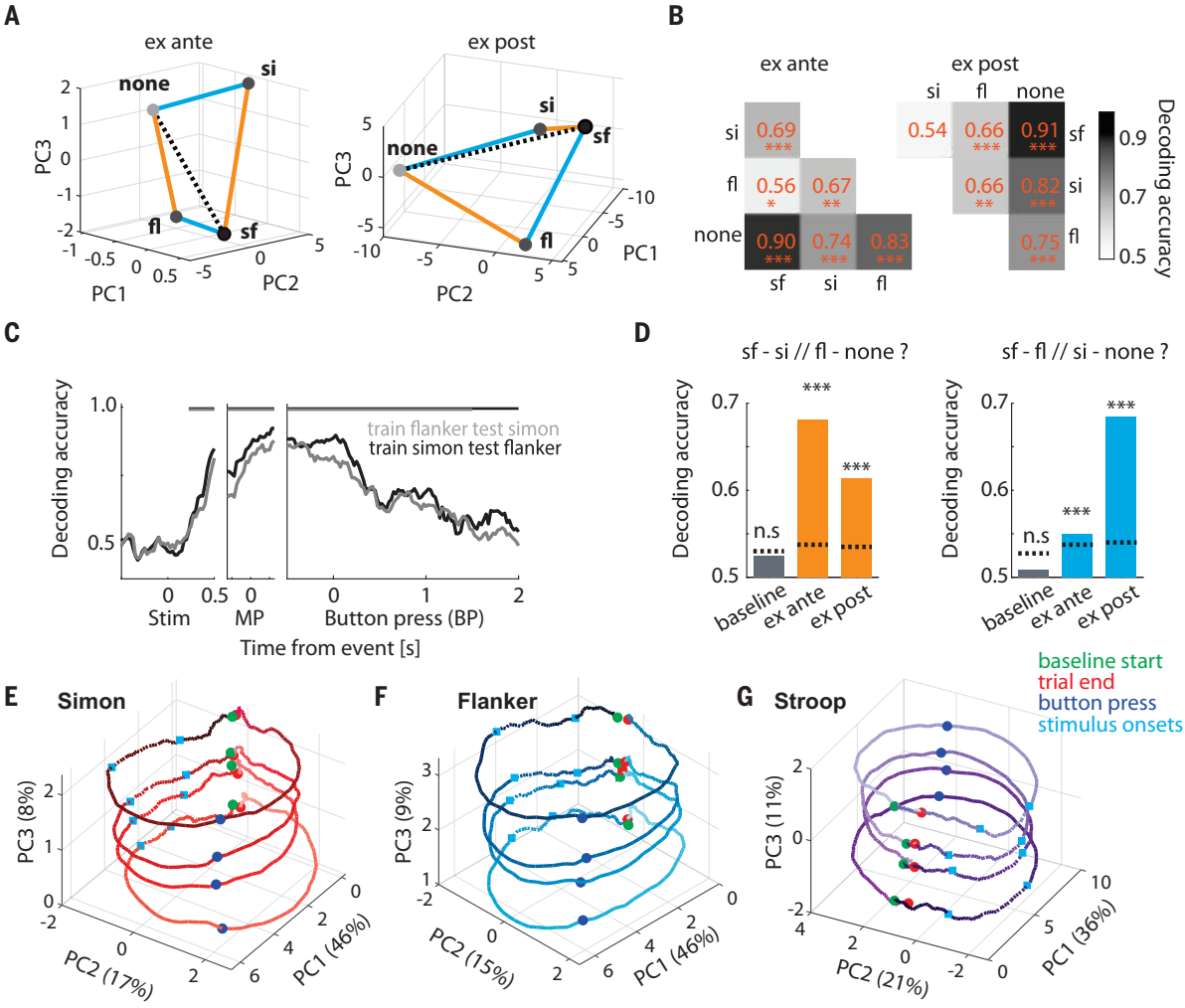


Fig. 4. State-space representation of conflict and conflict probability. (**A**) State-space representation of conflict (left, ex ante; right, ex post) in MSIT task visualized in PCA space. Dotted line is the vector used to classify pairs of conflict conditions in (B). (**B**) Decoding accuracy from classification of pairs of conflict conditions in MSIT. (**C**) Coding dimensions invariant between Simon and flanker conflict. At each time point, a decoder is trained on Simon versus non-Simon and tested on held-out flanker versus nonflanker trials (black) and vice versa (gray). (**D**) Conflict representations are compositional. Decoders trained on one edge of the parallelogram were able to differentiate between conditions along the opposite parallel edge [orange and blue edges shown in (A), respectively, shown on left and right]. Dotted lines show 97.5th percentile of the null distribution. (**E** to **G**) State-space representation of CP in MSIT and Stroop, visualized in PCA space. Green dots mark baseline start, the two cyan squares delineate the range of stimulus onsets, blue dots mark button press, and red dots mark end of trial. Trials are aligned to button press. Color fades as the trial progresses. Numbers signify percentage of variance explained by each principal component (PC). *P < 0.05; **P < 0.01; ***P < 0.001; n.s., not significant (P > 0.05). MP, midpoint; BP, button press; Stim, stimulus onset.

able to decode the two classes connected by the opposite edge (and vice versa). Indeed, a decoder trained to differentiate sf from fl trials, which is simply the vector connecting sf and fl (blue edge in Fig. 4A), was able to decode si from nonconflict trials projected to this axis with above-chance performance and vice versa (Fig. 4D, P < 0.001 for both the ex ante and ex post data, permutation test). The same was true for the other pair of edges (Fig. 4D, testing blue edges in Fig. 4A; P < 0.001 for both the ex ante and ex post data, permutation test). The parallelism was imperfect because the decoding accuracy, while well above chance, was relatively low (<70%) compared with within-condition decoding performance (Fig. 4C). Neurons that demonstrated nonlinear mixed selectivity for Simon and flanker conflicts (as measured by the F statistic of the interaction term between Simon and flanker derived from an analysis of variance model) contributed the most to this deviation from parallelism at the population level [fig. S9F, correlation coefficient (r) = 0.74, P < 0.001, for ex post data; fig. S9E, r = 0.75, P < 0.001, for ex ante data; Spearman's rank correlation). This representation structure was disrupted on error trials: generalization performance dropped in the ex ante (fig. S9D; for both edges, 68 and 61% on correct trials versus 56 and 47% on error trials) as well as the ex post epoch (fig. S9D; for both edges, 55 and 68% on correct trials versus 51 and 59% on error trials) on error trials.

Representational geometry of conflict probability

Conflict probability representation can be viewed as a state (an initial condition) that is present before stimulus onset and to which the population returns after completing a trial. To test this idea, we binned trials by CP quartiles (four "levels" were generated as trial labels) for each session and aggregated neuronal data across different sessions to generate pseudopopulations. Principal components analysis (PCA), performed separately for each type of conflict, revealed that the variability across

different CP levels (8 to 11% of explained variance) was captured mostly by a single axis (PC3s in Fig. 4, E to G; green dots mark trial start, red dots mark trial end), which was orthogonal to time-dependent but condition-independent firing rate changes (captured by PC1s and PC2s, 57 to 63% of explained variance). We quantified this impression in the full neural state space (not in PCA space). Before stimulus onset (demarcated by green and cyan filled circles in Fig. 4, E to G), the neural state evolved with low speed (blue points in fig. S9, G to I). The neural state then changed more quickly after stimulus onset (orange points in fig. S9, G to I; $P < 0.001$, paired t test), before eventually returning to a low-speed state that hovered near the starting position (red filled circles in Fig. 4, E to G; yellow points in fig. S9, G to I). The distance between the four trajectories was also kept stable across time. A coding dimension in the full neural space (which corresponds to PC3 in Fig. 4, E and F) was constructed by connecting the means of neural representations of the first and fourth CP levels. The held-out single-trial data, when projected onto this coding dimension, separated into four levels with their ordering (in projection values) being strongly correlated with the ordering of CP levels (fig. S9, J to L). This ordering is maintained into the baseline period of the next trial (fig. S9, J to L). For single sessions, CP predicted using spike counts from all MFC neurons in the session as regressors were correlated with the true values of CP trial by trial (fig. S15, M and N). These analyses in the full neural space confirmed the insights gained from PCA visualization in Fig. 4, E to G.

Domain-general performance-monitoring signals at the population level

We next investigated whether the geometry of performance-monitoring representations supported readouts that are invariant across MSIT and Stroop, while simultaneously allowing robust separation of conditions specific to MSIT. Within MSIT, a geometry can be extracted that supports invariance across types of conflicts while keeping the four conflict conditions separate (Fig. 4, A to C). However, it is unclear whether such representation is specific to a single task and requires construction of a task set. We thus studied the activity of the same neurons in two behavioral tasks performed separately (table S1 shows session information). We used demixed PCA (dPCA) (*57*) to factorize population activity into coding dimensions for performance monitoring variables (error, conflict, and conflict probability) and for the task identity. The dPCA provided a principled way to optimize for cross-task decoding and could serve as an existence proof for domain generality. The statistical significance of the dPCA coding dimensions was similarly assessed by out-of-sample decoding. Successful "demixing" should correspond to (near) orthogonality between the performance monitoring and task identity coding dimension.

The dPCA coding dimensions extracted using error and conflict contrasts (Stroop-Simon and Stroop-flanker separately) each explained 10 to 21% of variance, supported task-invariant decoding across time, and were orthogonal to the task identity dimension (see Fig. 5A for error, Fig. 5B for Stroop-Simon, fig. S10A for Stroop-flanker conflict decoding over time, and Fig. 5C for conflict decoding restricted in the ex ante and ex post epochs separately; for all clusters, $P < 0.01$, cluster-based permutation tests; clusters with significant decoding performance demarcated by horizontal bars; for error, angle = 92.1°, $P = 0.92$, $\tau = 0.005$; for Stroop-Simon conflict, angle = 81.9°, $P = 0.31$, $\tau = 0.04$; for Stroop-flanker conflict, angle = 80.8°, $P = 0.09$, $\tau = 0.06$; Kendall's rank correlation).

The task-invariant conflict coding dimension (extracted by removing the task difference between MSIT sf and Stroop conflict trials and between nonconflict trials in both tasks) was also able to differentiate five out of six pairs of conflict conditions with high accuracy (60 to 90%) within MSIT in both the ex ante and ex post epochs (Fig. 5C; for statistics, see figure, permutation tests). Similar task-invariant decoding performance was obtained when restricting to only trials with similar RTs across conditions, suggesting again that task-invariant representations of error and conflict were not due to subjective difficulty (fig. S10, B to E), for which RT is a proxy (*56*).

Similarly, the coding dimensions for Stroop-Simon conflict probability (Fig. 5D and fig. S10G) and Stroop-flanker conflict probability (fig. S10, F and H) each explained 27 to 35% of variance, supported task-invariant decoding, and were orthogonal to the task identity dimension ($P > 0.05$, Kendall's rank correlation). Notably, this task generalizability did not compromise the capacity of this coding dimension to separate different levels of CP within Stroop or MSIT (Fig. 5D and fig. S10, F to H).

As a control analysis, we also examined whether a decoder trained on a single task would generalize to the other task. These "within-task" coding dimensions supported decoding of data from the other task with high accuracy, albeit with slightly lower performance than the dPCA coding dimensions that were optimized for cross-task decoding (fig. S12, A to F, for decoding performance, weight correlation between tasks). This is expected given the robust correlation between the weights of these within-task coding dimensions (scatters in fig. S12, G to M). This correlation was confirmed by the fact that the angles between the two within-task coding dimensions were significantly smaller than the null distribution (which approximately equaled 90°; inset histograms in fig. S12, G to M). The angle between the task-invariant coding dimensions optimized by dPCA and the within-task dimensions trained with either Stroop or MSIT data alone was similar (angularly "equidistant"; angles reported in insets, fig. S12, G to M). The optimization performed by dPCA can thus be understood geometrically: It sought a coding dimension that is angularly "centered" between each of the within-task dimensions trained using data from one task only. Task-invariant representations for all performance monitoring variables can be found using simultaneously recorded data at the single-session level in subjects with enough neurons (fig. S15), highlighting that it is the same population of neurons that subserves such a geometry.

Subjects responded more slowly on the trials that followed error trials in both tasks, demonstrating robust post-error slowing (PES: 40 ± 1 ms and 44 ± 1 ms in Stroop and MSIT, respectively; fig. S8D). However, this trial-by-trial measure of PES was not significantly correlated with spike rates (on the error trials) of error neurons in either task individually, or for error neurons that signaled errors in both tasks (see methods for definition of "task-invariant"; fig. S8E).

Domain-general performance-monitoring signals at the single-neuron level

We sought to understand how tuning profiles of single neurons contributed to the population geometry that simultaneously allowed task-invariant and task-specific readouts. While some neurons encoded error, conflict, and CP in a task-invariant manner (Figs. 2 and 3), others encoded these variables in only one task. The extent to which single-neuron tuning depended on the task was assessed using partial correlations ("task-invariant turning strength"; higher correlation with variable of interest after controlling for task variable means higher task invariance). According to this test, neurons that had significant correlation were labeled "task-invariant" (65 to 83%), and those with insignificant cross-task correlation but significant correlation within either task were labeled "task-dependent" (17 to 35%; Fig. 5, E and F, and fig. S13, red and cyan slices, respectively). Across all neurons, the dPCA weight assigned to each neuron was strongly correlated with the neuron's task-invariant tuning strength (scatter plots in Fig. 5, E and F, and fig. S13; $P < 0.001$ for all panels, Spearman rank correlation). Both the task-invariant and the task-dependent neurons were assigned significant larger weights compared with nonselective ("other") neurons (Fig. 5, E and F, and fig. S13, dot density plots). Similar conclusions hold when using a

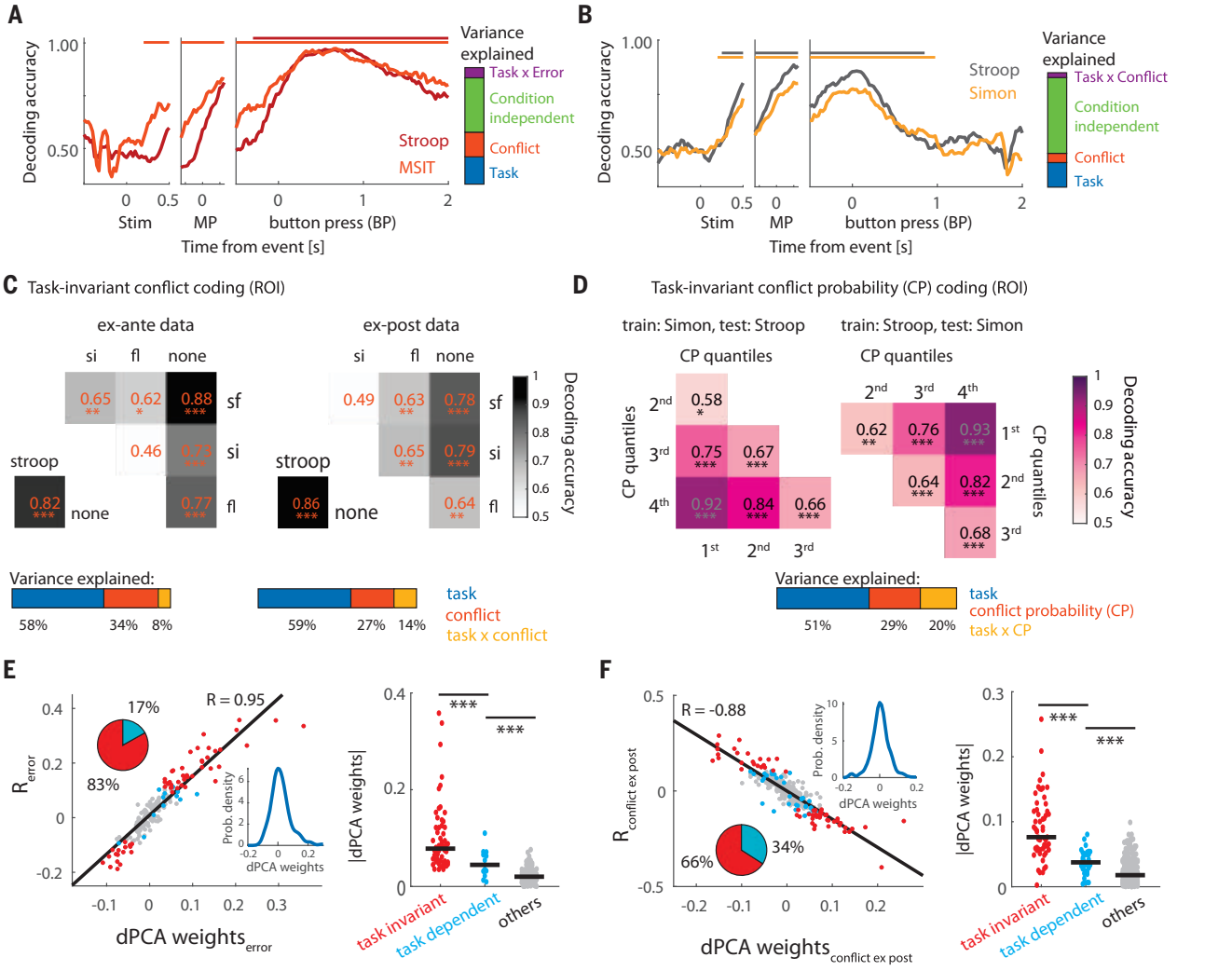


Fig. 5. Domain-general representation of performance-monitoring signals. (**A**) Task-invariant decoding of errors. Bar on the right shows the variance explained by the different dPCA components. Data from the whole trial were used. (**B**) Task-invariant decoding of conflict. The bar on the right represents variance explained by the different dPCA components (see figure legend for color code). Data from the whole trial were used. (**C**) Separability of conflict conditions along the domain-general conflict axis in both the ex ante (left) and ex post epochs (right). The dPCA coding dimension was constructed (separately for each epoch) using Stroop, sf conflict, and nonconflict trials and supported decoding of Stroop, Simon, and flanker conflicts ("task-invariant") as well as separation of Simon and flanker conditions. (**D**) Task-invariant decoding of CP. The dPCA coding axis was constructed using Stroop and Simon CP (binned by quartiles into four levels) and supported pairwise decoding of CP levels in both tasks. For (C) and (D), horizontal colored bars at the bottom show variance explained of dPCA coding dimensions. Data from single regions of interest (ROIs) were used. (**E** and **F**) Relationship between task-invariant single-neuron tuning strength of error (E), conflict ex post (F), and the corresponding dPCA weights. Pie charts show the percentages of task-invariant neurons (red slice) that had a significant main effect for the performance monitoring variable and those that had significant effect only in either MSIT or Stroop ("task-dependent", blue slice). Scatter plots (left) show significant correlation between task-invariant coding strength and the corresponding dPCA weights. The *y* axis shows correlation of firing rate of a neuron with the given variable, after removing task information by partial correlation (see methods). *P < 0.05; **P < 0.01; ***P ≤ 0.001; n.s., not significant (P > 0.05).

permutation-based feature-importance measure instead of the dPCA weights (fig. S14; see methods for details). Notably, the distributions of dPCA weights or feature importance (probability density shown in insets in Fig. 5, E and F, and figs. S13 and S14) across neurons, with many showing nonzero weights, suggested that the task-invariant coding was not driven by few neurons but instead reflected population activity.

Neurons with higher self-similarity for baseline spike counts contributed more to the task-invariant coding dimension of conflict probability after controlling for single-neuron coding strength (partial correlation between absolute values of dPCA weights and DFA α values; r = 0.16, P = 0.003 for Stroop-Simon conflict probability; r = 0.20, P < 0.001 for Stroop-flanker conflict probability).

Comparison between the dACC and the pre-SMA

The distributions of neuronal selectivity were similar in dACC and pre-SMA (fig. S3, A and B). Both areas contained similar proportions of task-invariant and task-dependent neurons. Both areas independently supported compositional conflict coding and domain-general readouts for error (fig. S11, A and B), conflict (fig. S11, C to F), and conflict probability (fig. S10, G to J). However, there were three notable differences between the two areas. First, the temporal profiles of decoding performance for error, conflict, and conflict probability were similar between the areas, but decoding accuracy in pre-SMA was consistently higher (fig. S5, A to E). Second, the task-invariant error

response appeared earlier in pre-SMA than in dACC ($\Delta_{latency}$ = 0.55 s for Stroop and 0.5 s for MSIT), consistent with our previous report in the Stroop task (*33*). Third, ex ante conflict information was first available in dACC, followed by pre-SMA (Fig. 3E; median difference = 138 ms; $P < 0.001$, Wilcoxon rank sum test; single-trial spike train latency). Third, by contrast, ex post conflict information was available first in pre-SMA, followed by dACC (Fig. 3E; median difference = 161 ms; $P = 0.002$, Wilcoxon rank sum test).

Discussion

Within the MFC, some neurons encoded error, conflict, and conflict probability in a task-invariant way; some encoded these variables exclusively in one task; and others multiplexed domain and task information to varying degrees. They were intermixed at similar anatomical locations and demonstrated nonlinear "mixed selectivity" (*47*, *48*) for task identity and performance monitoring variables. Such complexity at the single-neuron level precludes a simplistic interpretation of domain generality. By contrast, population activity can be factorized into a task dimension and a task-invariant dimension that were, in most cases, orthogonal to each other. The geometry of MFC population representation allows simple linear decoders to read out performance monitoring variables with high accuracy (>80%) in both tasks, and simultaneously to differentiate different types of conflict or different levels of estimated conflict probability within a task. Notably, it is the same group of neurons that gave rise to this geometry. This finding contrasts with neuroimaging studies that either report domain-specific and domain-general conflict signals encoded by distinct groups of voxels (*58*, *59*) or an absence of domain-general conflict signals (*60*) within the MFC.

Several studies have proposed that the lateral prefrontal cortex (LPFC) is topographically organized to subserve cognitive control, with more abstract processing engaging the anterior regions (*1*–*3*, *61*, *62*). Unlike in the LPFC, domain-general and domain-specific neurons were intermixed within the MFC. The representational geometry we report here is well suited to provide performance-monitoring signals to these subregions: Downstream neurons within these LPFC regions can select performance-monitoring information at different degrees of abstraction by adjusting connection weights, similar to input selection mechanisms described in the PFC (*63*). Curiously, we did not observe a prominent difference between the pre-SMA and the dACC in the degree of domain generality of performance monitoring, which may be related to the fact that these two areas are highly interconnected.

Domain-general error signals

A key component of performance monitoring and metacognitive judgment is the ability to self-monitor errors without relying on external feedback (*4*, *37*, *40*, *41*, *64*, *65*). A subset of neurons signaled error not only in the Stroop task [as previously reported in (*33*)] but also in MSIT. The error signal is thus domain-general, that is, abstracted away from the sensory and motor details as well as the types of response conflicts encountered across the two tasks (Fig. 2B). At the population level, these domain-general error neurons enabled trial-by-trial readout of self-monitored errors with >90% accuracy equal across tasks (Fig. 5, A and E). Compatible with earlier results (*33*), the activity of neither task-specific nor task-general error neurons directly predicted the extent of PES. This result is consistent with the evaluative roles of MFC error signals and suggests that the control process may lie outside of the MFC. Given the fMRI-BOLD finding that the MFC is a domain-general substrate for metacognition (*66*), an interesting open question is whether the same neural mechanisms we describe here support metacognitive judgment across different domains, such as perceptual or memory confidence.

Domain-specific performance monitoring

The causes of conflict and errors in the two tasks differ: distraction by the prepotent tendency to read in the Stroop task, and distraction by location of target number (Simon) or by numbers flanking the target (flanker) in the MSIT. Such performance perturbations call for specific compensatory mechanisms, such as suppressing attention to task-irrelevant stimulus dimensions. Consistent with this requirement, a subset of neurons signaled errors, conflict, and conflict probability exclusively in one task, giving rise to a task identity dimension that supported robust decoding of which task the performance disturbance occurred in (>90% accuracy). This provides the domain-specific information about the sources of performance disturbances for cognitive control, consistent with the reported role of MFC neurons in credit assignment (*16*). The existence of task-specific neurons also suggests that the performance monitoring circuitry can be rapidly and flexibly reconfigured in different tasks to subserve different task sets (*67*, *68*), consistent with the rapid reconfiguration of functional connectivity among cognitive control networks to enable novel task performance (*69*).

Compositionality of conflict representation

Compositionality of conflict representation can be formulated as a problem of generalization: If Simon and flanker conflict are linearly additive, decoders trained to identify the presence of only Simon or flanker conflict should continue to do so when both types of conflict are present. We found that this was the case, with the neural state approximately equal to the vector sum of the two neural states when the two types of conflict are present individually. The (approximate) factorization of conflict representation is important for both domain-specific and domain-general adaptation; when different types of conflict occur simultaneously and the representation can be factorized, downstream processes responsible for resolving each type of conflict can all be initiated. On the other hand, domain-general processes can also read out the representation as a sum and initiate domain-general adaptations.

Estimating control demand enabled by ex post conflict neurons

Our model for conflict probability estimation predicts that conflict should be signaled twice: once during response competition (ex ante) and again after the action has been committed (ex post). Whereas the former, predicted by conflict monitoring theory (*13*, *70*), provides transitory metacognitive evidence of conflict and is important for recruiting within-trial cognitive control (*70*), the latter provides a stable "indicator" of whether conflict occurs or not. Both signals arise independently of external feedback, thus qualifying as correlates of metacognitive self-monitoring (*64*, *65*). One interpretation for the ex post signals (conflict or error), as proposed by connectionist models of conflict monitoring, is that they reflect conflict between the committed response and continuing stimulus processing (*70*, *71*), which should also activate the ex ante conflict neurons. However, our data demonstrated that these two types of conflict signals were encoded by separate neurons and their multivariate coding patterns differed substantially enough to prevent generalization (fig. S6, C to E).

There is significant overlap between error neurons and ex post conflict neurons. Common coding dimensions that simultaneously decode both error and conflict do exist, although the decoding accuracy for conflict is significantly lower than for error (fig. S9, A and B). In macaque SEF, ex post activation is found after noncanceled and successfully canceled saccades (*41*). These ex post evaluative signals may reflect a common process that compares a corollary discharge signal (conveying the actual state of action selection and cognitive control) with cognitive control state predicted by "forward models" (*72*–*74*) and may underlie sense of agency (*75*). Future work is needed to test this hypothesis.

The neurons that reported conflict probability changed their firing rates trial by trial, and this "updating" in firing rates occurred primarily upon commitment of an action.

The distinct functional properties of conflict probability neurons, that is, significantly narrower extracellular spike waveforms and higher self-similarity of baseline spike counts (fig. S4, C to H), suggest that they may be interneurons with strong recurrent connectivity, consistent with a prior report where dACC neurons that encode past outcomes have narrower extracellular waveforms (*76*). At the population level, these single-neuron properties contribute to the formation of task-invariant line attractor dynamics. Computational modeling demonstrates the importance of inhibitory interneurons in maintaining information in working memory, which occurs on the scale of several seconds (*77*). Similar circuit-level mechanisms could provide a basis for retaining the history of performance monitoring and reward, which is on the scale of several minutes. These conflict probability neurons might thus provide a neural substrate for proactive control and learning the value of control, both of which require stable maintenance of learned information (*13*, *51*). The trade-off between flexible updating and stable maintenance of performance monitoring information remains an open question.

Materials and methods summary

Detailed materials and methods with references can be found in the supplementary materials. Briefly, the subjects were 34 patients who were undergoing intracranial monitoring of epileptic seizures using hybrid depth electrodes with embedded microwires. Spikes were detected and sorted using a template-matching algorithm, and only well-isolated single neurons were analyzed. Electrodes were localized using postoperative imaging, and all included subjects had well-isolated neuronal activity on at least one dACC or pre-SMA electrode.

Subjects performed speeded version of the Stroop and MSIT tasks. For the Stroop task, subjects were instructed to name the color in which the word shown on the screen (red, green, or blue) was printed (red, green, or blue). For the MSIT task, subjects were shown three numbers (0, 1, 2, and/or 3), out of which two were the same and the third was different (target), and were instructed to report the target number identity. After each response, the stimulus screen was immediately blanked (1 s), followed by a feedback screen ("correct," "incorrect," or "too slow"; 1 s). Trial sequences were randomized.

We used a hierarchical Bayesian model, which iteratively updates an internal estimate of conflict probability trial by trial. Parameters were the learning rate α and conflict probability (q_s for Stroop, q_{si} for Simon, and q_{fl} for flanker conflicts). Data used were: (i) trial congruency o (indicator function; o_s for Stroop, o_{si} for Simon, and o_{fl} for flanker congruency), assumed to be generated by the conflict probability; (ii) reaction time RT, assumed to be generated by a drift-diffusion model with the correct and wrong choice at the two bounds and the midpoint as the starting point. The statistical significance of the drift rates and drift rate bias coefficients was determined directly by comparing the posterior distribution of the group-level parameters to each other or to zero.

Neurons were selected using a Spearman's rank correlation between spike counts and a given performance monitoring variable ($P < 0.05$; error versus correct, conflict level, and conflict probability). Epochs used for selection were the baseline period (from 1.5 s before stimulus onset to stimulus onset), the ex ante epoch (a 500-ms window centered on the midpoint between stimulus onset and button press), and the ex post period (1-s window start at button press).

For decoding, one trial for each neuron from one condition was randomly selected and concatenated (across neurons) to form a testing data matrix. The rest of the trials were averaged for each condition and concatenated to form a training data matrix ("training means"). For within-task analysis, this matrix was used to compute the coding dimensions (by subtraction between conditions). For cross-task generalization analyses, training data matrices from both tasks were concatenated, and demixed PCA (*57*) was used to extract the task-invariant coding dimensions (except in fig. S12, in which cross-task decoding was evaluated by training in each task separately). Both testing and training data were then projected onto the identified coding dimensions. The predicted labels for the testing data were assigned according to the label of the nearest neighbor of the training data. This decoding procedure was repeated 1000 times (resulting in 1000 single-trial testing data matrices and the corresponding training data matrices). To determine statistical significance, we permuted the trial labels 500 times, and for each permutation, we repeated all above steps to generate a null distribution.

To assess the compositionality of conflict representation by decoding (Fig. 4D), the following coding dimensions were used: one flanked by nonconflict and si training means (Fig. 4A, blue), one flanked by fl and sf training means (blue), one flanked by fl and nonconflict trial averages (orange), and one flanked by si and sf training means (orange). Held-out testing data from conditions flanking one of the blue (or orange) pairs of edges were then projected to the opposite edge in the pair and classified by the training data defining this edge.

REFERENCES AND NOTES

1. E. Koechlin, C. Ody, F. Kouneiher, The architecture of cognitive control in the human prefrontal cortex. *Science* **302**, 1181–1185 (2003). doi: 10.1126/science.1088545; pmid: 14615530
2. D. Badre, M. D'Esposito, Functional magnetic resonance imaging evidence for a hierarchical organization of the prefrontal cortex. *J. Cogn. Neurosci.* **19**, 2082–2099 (2007). doi: 10.1162/jocn.2007.19.12.2082; pmid: 17892391
3. D. Badre, D. E. Nee, Frontal cortex and the hierarchical control of behavior. *Trends Cogn. Sci.* **22**, 170–188 (2018). doi: 10.1016/j.tics.2017.11.005; pmid: 29229206
4. M. Ullsperger, C. Danielmeier, G. Jocham, Neurophysiology of performance monitoring and adaptive behavior. *Physiol. Rev.* **94**, 35–79 (2014). doi: 10.1152/physrev.00041.2012; pmid: 24382883
5. R. E. Passingham, S. L. Bengtsson, H. C. Lau, Medial frontal cortex: From self-generated action to reflection on one's own performance. *Trends Cogn. Sci.* **14**, 16–21 (2010). doi: 10.1016/j.tics.2009.11.001; pmid: 19969501
6. J. D. Cosman, K. A. Lowe, W. Zinke, G. F. Woodman, J. D. Schall, Prefrontal control of visual distraction. *Curr. Biol.* **28**, 414–420.e3 (2018). doi: 10.1016/j.cub.2017.12.023; pmid: 29358071
7. C. Danielmeier, T. Eichele, B. U. Forstmann, M. Tittgemeyer, M. Ullsperger, Posterior medial frontal cortex activity predicts post-error adaptations in task-related visual and motor areas. *J. Neurosci.* **31**, 1780–1789 (2011). doi: 10.1523/JNEUROSCI.4299-10.2011; pmid: 21289188
8. T. Egner, J. Hirsch, Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. *Nat. Neurosci.* **8**, 1784–1790 (2005). doi: 10.1038/nn1594; pmid: 16286928
9. B. A. Purcell, R. Kiani, Neural mechanisms of post-error adjustments of decision policy in parietal cortex. *Neuron* **89**, 658–671 (2016). doi: 10.1016/j.neuron.2015.12.027; pmid: 26804992
10. J. G. Kerns *et al*., Anterior cingulate conflict monitoring and adjustments in control. *Science* **303**, 1023–1026 (2004). doi: 10.1126/science.1089910; pmid: 14963333
11. A. W. MacDonald 3rd, J. D. Cohen, V. A. Stenger, C. S. Carter, Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science* **288**, 1835–1838 (2000). doi: 10.1126/science.288.5472.1835; pmid: 10846167
12. E. K. Miller, J. D. Cohen, An integrative theory of prefrontal cortex function. *Annu. Rev. Neurosci.* **24**, 167–202 (2001). doi: 10.1146/annurev.neuro.24.1.167; pmid: 11283309
13. A. Shenhav, M. M. Botvinick, J. D. Cohen, The expected value of control: An integrative theory of anterior cingulate cortex function. *Neuron* **79**, 217–240 (2013). doi: 10.1016/j.neuron.2013.07.007; pmid: 23889930
14. R. Nigbur, M. X. Cohen, K. R. Ridderinkhof, B. Stürmer, Theta dynamics reveal domain-specific control over stimulus and response conflict. *J. Cogn. Neurosci.* **24**, 1264–1274 (2012). doi: 10.1162/jocn_a_00128; pmid: 21861681
15. S. D. McDougle *et al*., Credit assignment in movement-dependent reinforcement learning. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 6797–6802 (2016). doi: 10.1073/pnas.1523669113; pmid: 27247404
16. M. Sarafyazd, M. Jazayeri, Hierarchical reasoning by neural circuits in the frontal cortex. *Science* **364**, eaav8911 (2019). doi: 10.1126/science.aav8911; pmid: 31097640
17. D. Badre, A. D. Wagner, Selection, integration, and conflict monitoring: Assessing the nature and generality of prefrontal cognitive control mechanisms. *Neuron* **41**, 473–487 (2004). doi: 10.1016/S0896-6273(03)00851-1; pmid: 14766185
18. A. Genovesio, P. J. Brasted, A. R. Mitz, S. P. Wise, Prefrontal cortex activity related to abstract response strategies. *Neuron* **47**, 307–320 (2005). doi: 10.1016/j.neuron.2005.06.006; pmid: 16039571
19. P. Domenech, S. Rheims, E. Koechlin, Neural mechanisms resolving exploitation-exploration dilemmas in the medial prefrontal cortex. *Science* **369**, eabb0184 (2020). doi: 10.1126/science.abb0184; pmid: 32855307
20. J. R. Wessel, A. R. Aron, On the globality of motor suppression: Unexpected events and their influence on behavior and cognition. *Neuron* **93**, 259–280 (2017). doi: 10.1016/j.neuron.2016.12.013; pmid: 28103476
21. Y. Niv, N. D. Daw, D. Joel, P. Dayan, Tonic dopamine: Opportunity costs and the control of response vigor. *Psychopharmacology* **191**, 507–520 (2007). doi: 10.1007/s00213-006-0502-4; pmid: 17031711
22. J. F. Cavanagh *et al*., Subthalamic nucleus stimulation reverses mediofrontal influence over decision threshold. *Nat. Neurosci.* **14**, 1462–1467 (2011). doi: 10.1038/nn.2925; pmid: 21946325
23. E. A. Crone, R. J. M. Somsen, B. Van Beek, M. W. Van Der Molen, Heart rate and skin conductance analysis of antecendents

and consequences of decision making. *Psychophysiology* **41**, 531–540 (2004). doi: 10.1111/j.1469-8986.2004.00197.x; pmid: 15189476
24. R. B. Ebitz, M. L. Platt, Neuronal activity in primate dorsal anterior cingulate cortex signals task conflict and predicts adjustments in pupil-linked arousal. *Neuron* **85**, 628–640 (2015). doi: 10.1016/j.neuron.2014.12.053; pmid: 25654259
25. P. R. Murphy, E. Boonstra, S. Nieuwenhuis, Global gain modulation generates time-dependent urgency during perceptual choice in humans. *Nat. Commun.* **7**, 13526 (2016). doi: 10.1038/ncomms13526; pmid: 27882927
26. E. Eldar, R. B. Rutledge, R. J. Dolan, Y. Niv, Mood as representation of momentum. *Trends Cogn. Sci.* **20**, 15–24 (2016). doi: 10.1016/j.tics.2015.07.010; pmid: 26545853
27. R. B. Rutledge, N. Skandali, P. Dayan, R. J. Dolan, A computational and neural model of momentary subjective well-being. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 12252–12257 (2014). doi: 10.1073/pnas.1407535111; pmid: 25092308
28. A. J. Shackman *et al*., The integration of negative affect, pain and cognitive control in the cingulate cortex. *Nat. Rev. Neurosci.* **12**, 154–167 (2011). doi: 10.1038/nrn2994; pmid: 21331082
29. T. Paus, Primate anterior cingulate cortex: Where motor control, drive and cognition interface. *Nat. Rev. Neurosci.* **2**, 417–424 (2001). doi: 10.1038/35077500; pmid: 11389475
30. P. Nachev, C. Kennard, M. Husain, Functional role of the supplementary and pre-supplementary motor areas. *Nat. Rev. Neurosci.* **9**, 856–869 (2008). doi: 10.1038/nrn2478; pmid: 18843271
31. F. Bonini *et al*., Action monitoring and medial frontal cortex: Leading role of supplementary motor area. *Science* **343**, 888–891 (2014). doi: 10.1126/science.1247412; pmid: 24558161
32. C. S. Carter *et al*., Anterior cingulate cortex, error detection, and the online monitoring of performance. *Science* **280**, 747–749 (1998). doi: 10.1126/science.280.5364.747; pmid: 9563953
33. Z. Fu *et al*., Single-neuron correlates of error monitoring and post-error adjustments in human medial frontal cortex. *Neuron* **101**, 165–177.e5 (2019). doi: 10.1016/j.neuron.2018.11.016; pmid: 30528064
34. T. Gazit *et al*., The role of mPFC and MTL neurons in human choice under goal-conflict. *Nat. Commun.* **11**, 3192 (2020). doi: 10.1038/s41467-020-16908-z; pmid: 32581214
35. S. R. Heilbronner, B. Y. Hayden, Dorsal anterior cingulate cortex: A bottom-up view. *Annu. Rev. Neurosci.* **39**, 149–170 (2016). doi: 10.1146/annurev-neuro-070815-013952; pmid: 27090954
36. N. Kolling *et al*., Value, search, persistence and model updating in anterior cingulate cortex. *Nat. Neurosci.* **19**, 1280–1285 (2016). doi: 10.1038/nn.4382; pmid: 27669988
37. A. Sajad, D. C. Godlove, J. D. Schall, Cortical microcircuitry of performance monitoring. *Nat. Neurosci.* **22**, 265–274 (2019). doi: 10.1038/s41593-018-0309-8; pmid: 30643297
38. S. A. Sheth *et al*., Human dorsal anterior cingulate cortex neurons mediate ongoing behavioural adaptation. *Nature* **488**, 218–221 (2012). doi: 10.1038/nature11239; pmid: 22722841
39. H. Tang *et al*., Cascade of neural processing orchestrates cognitive control in human frontal cortex. *eLife* **5**, e12352 (2016). doi: 10.7554/eLife.12352; pmid: 26888070
40. S. Ito, V. Stuphorn, J. W. Brown, J. D. Schall, Performance monitoring by the anterior cingulate cortex during saccade countermanding. *Science* **302**, 120–122 (2003). doi: 10.1126/science.1087847; pmid: 14526085
41. V. Stuphorn, T. L. Taylor, J. D. Schall, Performance monitoring by the supplementary eye field. *Nature* **408**, 857–860 (2000). doi: 10.1038/35048576; pmid: 11130724
42. C. Wang, I. Ulbert, D. L. Schomer, K. Marinkovic, E. Halgren, Responses of human anterior cingulate cortex microdomains to error detection, conflict monitoring, stimulus-response mapping, familiarity, and orienting. *J. Neurosci.* **25**, 604–613 (2005). doi: 10.1523/JNEUROSCI.4151-04.2005; pmid: 15659596
43. Z. M. Williams, G. Bush, S. L. Rauch, G. R. Cosgrove, E. N. Eskandar, Human anterior cingulate neurons and the integration of monetary reward with motor responses. *Nat. Neurosci.* **7**, 1370–1375 (2004). doi: 10.1038/nn1354; pmid: 15558064
44. S. Bernardi *et al*., The geometry of abstraction in the hippocampus and prefrontal cortex. *Cell* **183**, 954–967.e21 (2020). doi: 10.1016/j.cell.2020.09.031; pmid: 33058757
45. C. Stringer, M. Pachitariu, N. Steinmetz, M. Carandini, K. D. Harris, High-dimensional geometry of population responses in visual cortex. *Nature* **571**, 361–365 (2019). doi: 10.1038/s41586-019-1346-5; pmid: 31243367
46. J. J. DiCarlo, D. D. Cox, Untangling invariant object recognition. *Trends Cogn. Sci.* **11**, 333–341 (2007). doi: 10.1016/j.tics.2007.06.010; pmid: 17631409
47. S. Fusi, E. K. Miller, M. Rigotti, Why neurons mix: High dimensionality for higher cognition. *Curr. Opin. Neurobiol.* **37**, 66–74 (2016). doi: 10.1016/j.conb.2016.01.010; pmid: 26851755
48. M. Rigotti *et al*., The importance of mixed selectivity in complex cognitive tasks. *Nature* **497**, 585–590 (2013). doi: 10.1038/nature12160; pmid: 23685452
49. G. R. Yang, M. R. Joglekar, H. F. Song, W. T. Newsome, X.-J. Wang, Task representations in neural networks trained to perform many cognitive tasks. *Nat. Neurosci.* **22**, 297–306 (2019). doi: 10.1038/s41593-018-0310-2; pmid: 30643294
50. J. Minxha, R. Adolphs, S. Fusi, A. N. Mamelak, U. Rutishauser, Flexible recruitment of memory-based choice representations by the human medial frontal cortex. *Science* **368**, eaba3313 (2020). doi: 10.1126/science.aba3313; pmid: 32586990
51. T. S. Braver, The variable nature of cognitive control: A dual mechanisms framework. *Trends Cogn. Sci.* **16**, 106–113 (2012). doi: 10.1016/j.tics.2011.12.010; pmid: 22245618
52. J. Jiang, J. Beck, K. Heller, T. Egner, An insula-frontostriatal network mediates flexible cognitive control by adaptively predicting changing control demands. *Nat. Commun.* **6**, 8165 (2015). doi: 10.1038/ncomms9165; pmid: 26391305
53. J. Jiang, K. Heller, T. Egner, Bayesian modeling of flexible cognitive control. *Neurosci. Biobehav. Rev.* **46**, 30–43 (2014). doi: 10.1016/j.neubiorev.2014.06.001; pmid: 24929218
54. T. E. J. Behrens, M. W. Woolrich, M. E. Walton, M. F. S. Rushworth, Learning the value of information in an uncertain world. *Nat. Neurosci.* **10**, 1214–1221 (2007). doi: 10.1038/nn1954; pmid: 17676057
55. Four additional Bayesian models were considered for model comparison: (i) DDM only, with no CP estimation; (ii) DDM with drift rate biases depending on the posterior means of CP (inferred separately); (iii) DDM with drift rate biases depending on CP updated by a deterministic learning rule with a single learning rate; and (iv) DDM with drift rate biases depending on previous trial conflict. The full conflict estimation model outperformed these alternative models (see table S2 for measures of model comparison and table S3 for model weights). Comparison with the first model suggests that subjects used a flexible instead of fixed learning rate in estimating CP. Comparison with the second model suggests that not only the CP but also its uncertainty provides useful information for predicting RT. The comparison with the fourth model, consistent with prior work (*52*), demonstrates that participants incorporated conflict from multiple past trials.
56. G. Gratton, M. G. Coles, E. Donchin, Optimizing the use of information: Strategic control of activation of responses. *J. Exp. Psychol. Gen.* **121**, 480–506 (1992). doi: 10.1037/0096-3445.121.4.480; pmid: 1431740
57. D. Kobak *et al*., Demixed principal component analysis of neural population data. *eLife* **5**, e10989 (2016). doi: 10.7554/eLife.10989; pmid: 27067378
58. J. Jiang, T. Egner, Using neural pattern classifiers to quantify the modularity of conflict-control mechanisms in the human brain. *Cereb. Cortex* **24**, 1793–1805 (2014). doi: 10.1093/cercor/bht029; pmid: 23402762
59. T. A. Niendam *et al*., Meta-analytic evidence for a superordinate cognitive control network subserving diverse executive functions. *Cogn. Affect. Behav. Neurosci.* **12**, 241–268 (2012). doi: 10.3758/s13415-011-0083-5; pmid: 22282036
60. P. A. Kragel *et al*., Generalizable representations of pain, cognitive control, and negative emotion in medial frontal cortex. *Nat. Neurosci.* **21**, 283–289 (2018). doi: 10.1038/s41593-017-0051-7; pmid: 29292378
61. S. Tsujimoto, A. Genovesio, S. P. Wise, Frontal pole cortex: Encoding ends at the end of the endbrain. *Trends Cogn. Sci.* **15**, 169–176 (2011). doi: 10.1016/j.tics.2011.02.001; pmid: 21388858
62. F. A. Mansouri, E. Koechlin, M. G. P. Rosa, M. J. Buckley, Managing competing goals—a key role for the frontopolar cortex. *Nat. Rev. Neurosci.* **18**, 645–657 (2017). doi: 10.1038/nrn.2017.111; pmid: 28951610
63. V. Mante, D. Sussillo, K. V. Shenoy, W. T. Newsome, Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* **503**, 78–84 (2013). doi: 10.1038/nature12742; pmid: 24201281
64. N. Yeung, C. Summerfield, Metacognition in human decision-making: Confidence and error monitoring. *Philos. Trans. R. Soc. London B Biol. Sci.* **367**, 1310–1321 (2012). doi: 10.1098/rstb.2011.0416; pmid: 22492749
65. S. M. Fleming, R. J. Dolan, The neural basis of metacognitive ability. *Philos. Trans. R. Soc. London Ser. B* **367**, 1338–1349 (2012). doi: 10.1098/rstb.2011.0417; pmid: 22492751
66. J. Morales, H. Lau, S. M. Fleming, Domain-general and domain-specific patterns of activity supporting metacognition in human prefrontal cortex. *J. Neurosci.* **38**, 3534–3546 (2018). doi: 10.1523/JNEUROSCI.2360-17.2018; pmid: 29519851
67. N. U. F. Dosenbach *et al*., A core system for the implementation of task sets. *Neuron* **50**, 799–812 (2006). doi: 10.1016/j.neuron.2006.04.031; pmid: 16731517
68. M. F. S. Rushworth, M. E. Walton, S. W. Kennerley, D. M. Bannerman, Action sets and decisions in the medial frontal cortex. *Trends Cogn. Sci.* **8**, 410–417 (2004). doi: 10.1016/j.tics.2004.07.009; pmid: 15350242
69. M. W. Cole *et al*., Multi-task connectivity reveals flexible hubs for adaptive task control. *Nat. Neurosci.* **16**, 1348–1355 (2013). doi: 10.1038/nn.3470; pmid: 23892552
70. M. M. Botvinick, T. S. Braver, D. M. Barch, C. S. Carter, J. D. Cohen, Conflict monitoring and cognitive control. *Psychol. Rev.* **108**, 624–652 (2001). doi: 10.1037/0033-295X.108.3.624; pmid: 11488380
71. N. Yeung, M. M. Botvinick, J. D. Cohen, The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychol. Rev.* **111**, 931–959 (2004). doi: 10.1037/0033-295X.111.4.931; pmid: 15482068
72. C.-C. Lo, X.-J. Wang, Cortico-basal ganglia circuit mechanism for a decision threshold in reaction time tasks. *Nat. Neurosci.* **9**, 956–963 (2006). doi: 10.1038/nn1722; pmid: 16767089
73. D. Subramanian, A. Alers, M. A. Sommer, Corollary discharge for action and cognition. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **4**, 782–790 (2019). doi: 10.1016/j.bpsc.2019.05.010; pmid: 31351985
74. W. H. Alexander, J. W. Brown, Medial prefrontal cortex as an action-outcome predictor. *Nat. Neurosci.* **14**, 1338–1344 (2011). doi: 10.1038/nn.2921; pmid: 21926982
75. P. Haggard, Sense of agency in the human brain. *Nat. Rev. Neurosci.* **18**, 196–207 (2017). doi: 10.1038/nrn.2017.14; pmid: 28251993
76. T. Kawai, H. Yamada, N. Sato, M. Takada, M. Matsumoto, Preferential representation of past outcome information and future choice behavior by putative inhibitory interneurons rather than putative pyramidal neurons in the primate dorsal anterior cingulate cortex. *Cereb. Cortex* **29**, 2339–2352 (2019). doi: 10.1093/cercor/bhy103; pmid: 29722795
77. N. Brunel, X.-J. Wang, Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition. *J. Comput. Neurosci.* **11**, 63–85 (2001). doi: 10.1023/A:1011204814320; pmid: 11524578
78. Z. Fu *et al*., Data for: The geometry of domain-general performance monitoring in the human medial frontal cortex, Open Science Framework (2022); https://osf.io/42r9c/.

ACKNOWLEDGMENTS

We thank the members of the Adolphs and Rutishauser labs, L. J. Jin, and S. Dong for discussion. We thank all subjects and their families for their participation and the staff of the Cedars-Sinai Epilepsy Monitoring Unit for their support. **Funding:** This work was supported by NIMH (R01MH110831 to U.R.), the NIMH Conte Center (P50MH094258 to R.A. and U.R.), the National Science Foundation (CAREER Award BCS-1554105), the BRAIN Initiative through the NIH Office of the Director (U01NS117839), and the Simons Foundation Collaboration on the Global Brain (R.A.). **Author contributions:** Z.F., R.A., and U.R. designed the study. Z.F. and U.R. collected and analyzed the data and implemented analysis procedures. Z.F., U.R., A.N.M., and R.A. wrote the paper. D.B. acquired and analyzed the behavioral control data. J.M.C. and C.M.R. provided patient care and facilitated experiments. A.N.M. performed surgery. **Competing interest:** The authors have no competing interests to declare. **Data and materials availability:** Data needed to reproduce results have been deposited at Open Science Framework (OSF) (*78*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm9922
Materials and Methods
Figs. S1 to S15
Tables S1 to S3
References (*79*–*98*)
MDAR Reproducibility Checklist

Submitted 27 October 2021; accepted 1 April 2022
10.1126/science.abm9922

RESEARCH ARTICLE SUMMARY

CORONAVIRUS

Increased risk of SARS-CoV-2 reinfection associated with emergence of Omicron in South Africa

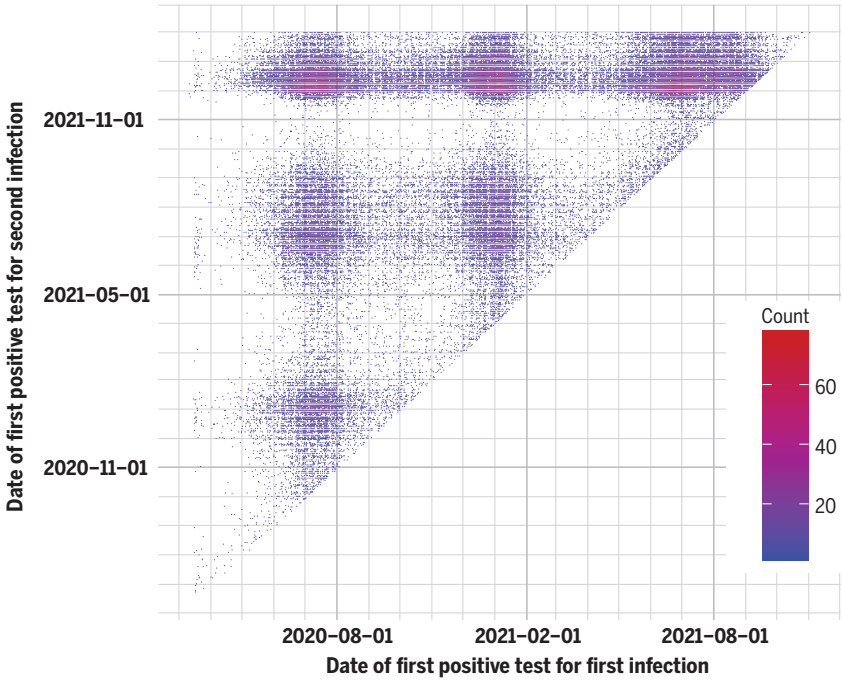
Juliet R. C. Pulliam*, Cari van Schalkwyk, Nevashan Govender, Anne von Gottberg, Cheryl Cohen, Michelle J. Groome, Jonathan Dushoff, Koleka Mlisana, Harry Moultrie

INTRODUCTION: Globally, there have been more than 404 million cases of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), with 5.8 million confirmed deaths as of February 2022. South Africa has experienced four waves of SARS-CoV-2 transmission, with the second, third, and fourth waves being driven by the Beta, Delta, and Omicron variants, respectively. A key question with the emergence of new variants is the extent to which they are able to reinfect those who have had a prior natural infection.

RATIONALE: We developed two approaches to monitor routine epidemiological surveillance data to determine whether SARS-CoV-2 reinfection risk has changed through time in South Africa in the context of the emergence of the Beta (B.1.351), Delta (B.1.617.2), and Omicron (B.1.1.529) variants. We analyzed line-list data on positive tests for SARS-CoV-2 with specimen receipt dates between 4 March 2020 and 31 January 2022 collected through South Africa's National Notifiable Medical Conditions Surveillance System. Individuals having sequential positive tests at least 90 days apart were considered to have suspected reinfections. Our routine monitoring of reinfection risk included comparison of reinfection rates with the expectation under a null model (approach 1) and estimation of the time-varying hazards of infection and reinfection throughout the epidemic (approach 2) based on model-based reconstruction of the susceptible populations eligible for primary and second infections.

RESULTS: A total of 105,323 suspected reinfections were identified among 2,942,248 individuals with laboratory-confirmed SARS-CoV-2 who had a positive test result at least 90 days before 31 January 2022. The number of reinfections observed through the end of the third wave in September 2021 was consistent with the null model of no change in reinfection risk (approach 1). Although increases in the hazard of primary infection were observed after the introduction of both the Beta and Delta variants, no corresponding increase was observed in the reinfection hazard (approach 2). Contrary to expectation, the estimated hazard ratio for reinfection versus primary infection was lower during waves driven by the Beta and Delta variants than for the first wave: the relative hazard ratio for wave 2 versus wave 1 was 0.71 [95% confidence interval (95% CI): 0.60 to 0.85]; the relative hazard ratio for wave 3 versus wave 1 was 0.54 (95% CI: 0.45 to 0.64). By contrast, the recent spread of the Omicron variant has been associated with an increase in reinfection hazard coefficient. The estimated relative hazard ratio for reinfection versus primary infection versus wave 1 was 1.75 (95% CI: 1.48 to 2.10) for the period of Omicron emergence (1 November 2021 to 30 November 2021) and 1.70 (95% CI: 1.44 to 2.04) for wave 4 versus wave 1. Individuals with identified reinfections since 1 November 2021 had experienced primary infections in all three prior waves, and an increase in third infections has been detected since mid-November 2021. Many individuals experiencing third infections had second infections during the third (Delta) wave that ended in September 2021, strongly suggesting that these infections resulted from immune evasion rather than waning immunity.

CONCLUSION: Population-level evidence suggests that the Omicron variant is associated with a marked ability to evade immunity from prior infection. In contrast, there is no population-wide epidemiological evidence of immune escape associated with the Beta or Delta variants. This finding has important implications for public health planning, particularly in countries such as South Africa with high rates of immunity from prior infection. The further development of methods to track reinfection risk during pathogen emergence, including refinements to assess the impact of waning immunity, account for vaccine-derived protection, and monitor the risk of multiple reinfections, will be important for future pandemic preparedness. ■



SARS-CoV-2 reinfection patterns in South Africa. South Africa has experienced four waves of SARS-CoV-2 transmission, each driven by the emergence of a new variant. Reinfection of previously infected individuals was relatively rare through the end of the third wave. Methods developed in South Africa to monitor reinfection trends led to the early detection of increased reinfection risk associated with the Omicron variant.

The list of author affiliations is available in the full article online.
*Corresponding author. Email: pulliam@sun.ac.za
This is an open-access article distributed under the terms of the Creative Commons Attribution license (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.
Cite this article as J. R. C. Pulliam *et al.*, *Science* **376**, eabn4947 (2022). DOI: 10.1126/science.abn4947

READ THE FULL ARTICLE AT
https://doi.org/10.1126/science.abn4947

RESEARCH ARTICLE

CORONAVIRUS

Increased risk of SARS-CoV-2 reinfection associated with emergence of Omicron in South Africa

Juliet R. C. Pulliam[1]*, Cari van Schalkwyk[1], Nevashan Govender[2], Anne von Gottberg[2,3], Cheryl Cohen[2,4], Michelle J. Groome[2,3], Jonathan Dushoff[1,5], Koleka Mlisana[6,7,8], Harry Moultrie[2,3]

We provide two methods for monitoring reinfection trends in routine surveillance data to identify signatures of changes in reinfection risk and apply these approaches to data from South Africa's severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) epidemic to date. Although we found no evidence of increased reinfection risk associated with circulation of the Beta (B.1.351) or Delta (B.1.617.2) variants, we did find clear, population-level evidence to suggest immune evasion by the Omicron (B.1.1.529) variant in previously infected individuals in South Africa. Reinfections occurring between 1 November 2021 and 31 January 2022 were detected in individuals infected in all three previous waves, and there has been an increase in the risk of having a third infection since mid-November 2021.

As of 31 January 2022, South Africa had >3.6 million cumulative laboratory-confirmed cases of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), which were concentrated in four waves of infection (Fig. 1). The first case was detected in early March 2020 and was followed by a wave that peaked in July 2020 and ended in September 2020. The second wave, which peaked in January 2021 and ended in February 2021, was driven by the Beta (B.1.351/501Y.V2/20H) variant, which was first detected in South Africa in October 2020 (*1*). The third wave, which peaked in July 2021 and ended in September 2021, was dominated by the Delta (B.1.617.2/478K.V1/21A) variant (*2*). In late November 2021, the Omicron (B.1.1.529/21K) variant was detected in Gauteng Province, the smallest yet most populous province in South Africa, and was associated with rapidly increasing case numbers (*3*). The estimated effective reproduction number in Gauteng based on polymerase chain reaction (PCR)–confirmed cases was 2.3 as of 18 November 2021, which was as high as had been seen at any point during the prior three waves, and peaked at >3 in late November 2021 (*4*, *5*). The proportion of positive PCR tests with S-gene target failure, a marker of the BA.1 sublineage of the Omicron variant, subsequently increased across all provinces (*6*).

After the emergence of three variants of concern (VOCs) in South Africa, a key question remaining in late 2021 was whether there was epidemiologic evidence of increased risk of SARS-CoV-2 reinfection with these variants (i.e., immune escape from natural infection). Laboratory-based studies suggest that convalescent serum has a reduced neutralizing effect on the Beta, Delta, and Omicron variants compared with wild-type virus in vitro (*7*–*12*); however, this finding does not necessarily translate into immune evasion at the population level.

To determine whether reinfection risk has changed through time, it is essential to account for potential confounding factors affecting the incidence of reinfection, namely, the changing force of infection experienced by all individuals in the population and the growing number of individuals eligible for reinfection through time. These factors are tightly linked to the timing of epidemic waves. We examined reinfection trends in South Africa using two approaches that account for these factors to address the question of whether circulation of VOCs has been associated with increased reinfection risk, as would be expected if their emergence were driven or facilitated by immune evasion.

Identification of and characterization of reinfections

We define a suspected reinfection as a positive SARS-CoV-2 test in an individual with at least one previous positive test whose most recent positive test occurred at least 90 days earlier. Based on routinely collected line-list data maintained by the National Institute for Communicable Diseases (NICD) with specimen receipt dates between 4 March 2020 and 31 January 2022, we identified 105,323 individuals with at least two suspected infections, 1778 individuals with at least three suspected infections, and 18 individuals with four suspected infections.

Time between successive positive tests

The distribution of times between successive positive tests for individuals' first and second infections has peaks near 170, 350, and 520 days (Fig. 2A). The shape of the distribution is strongly influenced by the timing of South Africa's epidemic waves, which have been spaced ~6 months apart. The first peak corresponds mainly to individuals whose primary infection and second infection occurred in consecutive waves (e.g., initially infected in wave 1 and reinfected in wave 2, initially infected in wave 2 and reinfected in wave 3, or initially infected in wave 3 and reinfected in wave 4), whereas the second peak corresponds mainly to individuals initially infected in wave 1 and reinfected in wave 3 or initially infected in wave 2 and reinfected in wave 4. The third peak corresponds to individuals initially infected in wave 1 and reinfected in wave 4.

Almost all suspected third infections occurred after 31 October 2021, i.e., during the period of Omicron circulation. The distribution of times between successive positive tests for individuals' second and third infections has peaks corresponding to those whose second infections occurred in the second and third waves.

Individuals with multiple suspected reinfections

A total of 1778 individuals who had three or more suspected infections were identified. Before the emergence of Omicron, most of these individuals initially tested positive during the first wave, with suspected reinfections associated with waves 2 and 3; however, 1492 individuals with multiple reinfections (83.9%) experienced their third infection after 31 October 2021, which suggests that most third infections were associated with transmission of the Omicron variant (Fig. 3).

Population-level reinfection trends in South Africa

The population at risk of reinfection has risen monotonically since the beginning of the epidemic, with relatively rapid increases associated with each wave (delayed by 90 days because of our definition of reinfection; Fig. 1B). No suspected reinfections were detected until 23 June 2020, after which the number of suspected reinfections increased gradually. The 7-day moving average of suspected second infections reached a peak of ~160 during the second epidemic wave and 350 during the third wave (Fig. 1). After the third wave, the number of reinfections began to increase markedly in mid-November 2021. During the fourth wave,

[1]South African DSI-NRF Centre of Excellence in Epidemiological Modelling and Analysis (SACEMA), Stellenbosch University, Stellenbosch, South Africa. [2]National Institute for Communicable Diseases, Division of the National Health Laboratory Service, Johannesburg, South Africa. [3]School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa. [4]School of Public Health, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa. [5]McMaster University, Hamilton, Ontario, Canada. [6]National Health Laboratory Service, Johannesburg, South Africa. [7]School of Laboratory Medicine and Medical Sciences, University of KwaZulu-Natal, Durban, South Africa. [8]Centre for the AIDS Programme of Research in South Africa (CAPRISA), Durban, South Africa.
*Corresponding author. Email: pulliam@sun.ac.za

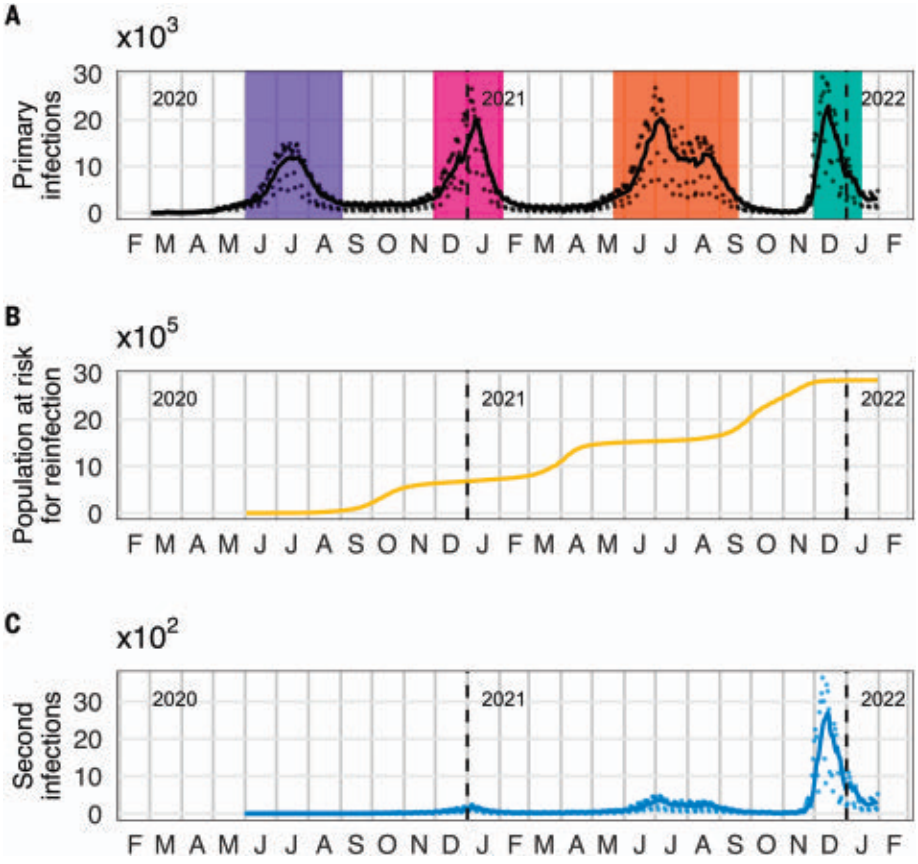


Fig. 1. Daily numbers of detected primary infections, individuals eligible to be considered for reinfection, and suspected reinfections in South Africa. (**A**) Time series of detected primary infections. Black line indicates 7-day moving average; black points are daily values. Colored bands represent wave periods, defined as the period for which the 7-day moving average of cases (detected infections and reinfections) was at least 15% of the corresponding wave peak (purple indicates wave 1; pink, wave 2; orange, wave 3; and turquoise, wave 4). (**B**) Population at risk for reinfection: individuals who tested positive at least 90 days ago and have not yet had a suspected reinfection. (**C**) Time series of suspected second infections. Blue line indicates 7-day moving average; blue points are daily values.

the 7-day moving average of suspected second infections reached nearly 2700, and the 7-day moving average of all suspected reinfections (including second, third, and fourth infections) reached ~2750.

Comparison of data with projections from a null model

We developed a catalytic model to project the expected number of reinfections through time under the assumption of a constant reinfection hazard coefficient (i.e., a null model of no change in reinfection risk). The model assumes that the reinfection hazard is proportional to the 7-day moving average of the total number of diagnosed infections (primary infections and reinfections). During our early monitoring of reinfection risk, we fitted the reinfection hazard coefficient to data from 2 June 2020 to 30 September 2020 to parameterize the null model of no change in the reinfection hazard coefficient through time and projected the number of reinfections through 30 June 2021. Based on this, we concluded that there was no population-level evidence of immune escape and recommended ongoing monitoring of reinfection trends (*13*).

Given that there was no evidence of divergence from the null projection during the second wave and to improve convergence of the Metropolis-Hastings Monte Carlo Markov Chain (MCMC) fitting algorithm, for the present analysis, we repeated the fitting process using the time window of 2 June 2020 to 28 February 2021 (representing the end of the month in which the second wave ended). This led to good convergence with regard to estimation of both the negative binomial dispersion parameter and the reinfection hazard coefficient (fig. S4) and allowed us to fit the model to all nine provinces. The 7-day moving average of observed reinfections and most individual daily values fell within the projection interval from the beginning of the projection period though the end of the third wave (Fig. 4). From early November 2021, however, the 7-day moving average of observed reinfections reached the upper bound of the projection interval, with many individual daily numbers falling well above the projection interval both nationally and in Gauteng (Fig. 4). This observed deviation from the projection under the null model is a signature of immune evasion, and the timing of this deviation suggests that it is associated with the emergence of the Omicron variant. A similar pattern has now been seen across all provinces in South Africa (figs. S5 to S7).

Estimation of time-varying infection and reinfection hazards

We also examined changes in the reinfection risk using a method that relies on reconstruction of the numbers of observed and unobserved first and second infections through time (see the materials and methods for details). On the basis of this approach, the estimated hazard coefficient for primary infection increased steadily through the end of the third wave, as expected under a combination of relaxing of restrictions, behavioral fatigue, and the introduction of variants with increased transmissibility (Beta and Delta). By contrast, the estimated hazard coefficient for reinfection remained relatively constant throughout this period, with the exception of an initial spike in mid-2020 (Fig. 5). Because both reinfection numbers and the population eligible for reinfection were very low at the time, this increase may be an artifact of intense follow-up of the earliest cases or simply noise caused by the small numbers. The estimated mean ratio of reinfection hazard to primary infection hazard decreased slightly from 0.15 in wave 1 to 0.12 in wave 2 and 0.09 in wave 3. The absolute values of the hazard coefficients and hazard ratio are sensitive to assumed observation probabilities for primary infections and reinfections; however, the temporal trends are robust (fig. S8).

The picture changed after the end of the third wave. Although there is substantial uncertainty in the estimated hazard coefficient for primary infection, it appeared to decrease from early October 2021, with a simultaneous increase in the estimated reinfection hazard coefficient (Fig. 5). This change became more marked from the beginning of November, with the mean ratio of reinfection hazard to primary infection hazard for the period from 1 November 2021 to the beginning of the fourth wave increasing to 0.25 and a mean ratio during the fourth wave of 0.27.

These findings are consistent with the estimates from a generalized linear mixed model based on the reconstructed dataset. In this analysis, the relative hazard ratio for wave 2 versus wave 1 was 0.71 [95% confidence interval (CI): 0.60 to 0.85] and for wave 3 versus wave 1 it was 0.54 (95% CI: 0.45 to 0.64). The

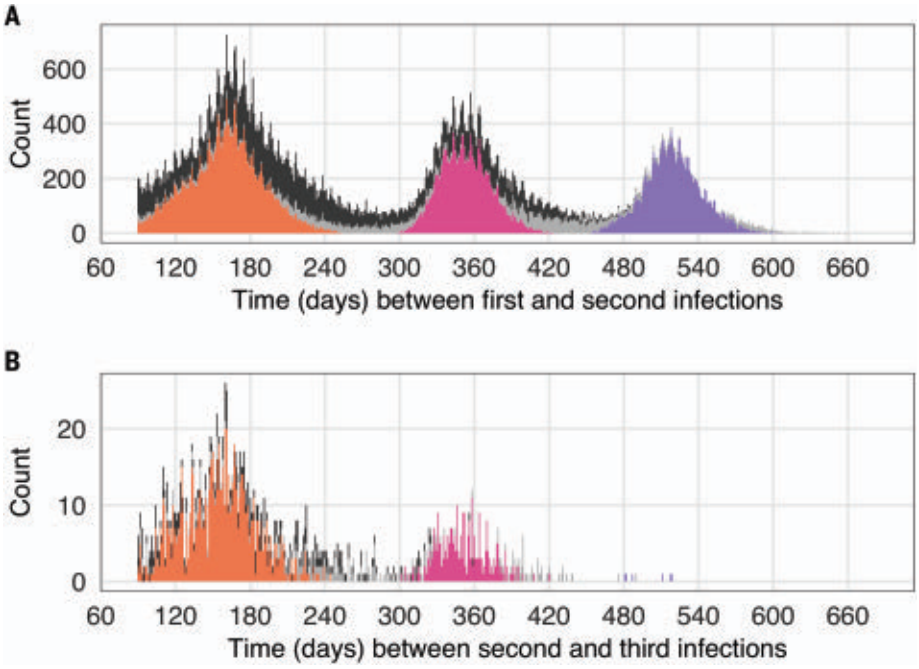


Fig. 2. Time between consecutive infections based on the time between successive positive tests. Note that the time since the previous positive test must be at least 90 days to be considered a reinfection. (**A**) Time in days between the last positive test of the putative first infection and the first positive test of the suspected second infection. (**B**) Time in days between the last positive test of the putative second infection and the first positive tests of the suspected third infection. Colors represent suspected reinfections diagnosed on or after 1 November 2021. In both panels, bars for these individuals are colored by the wave during which the previous infection occurred (purple indicates wave 1; pink, wave 2; orange, wave 3; and light gray, interwave).

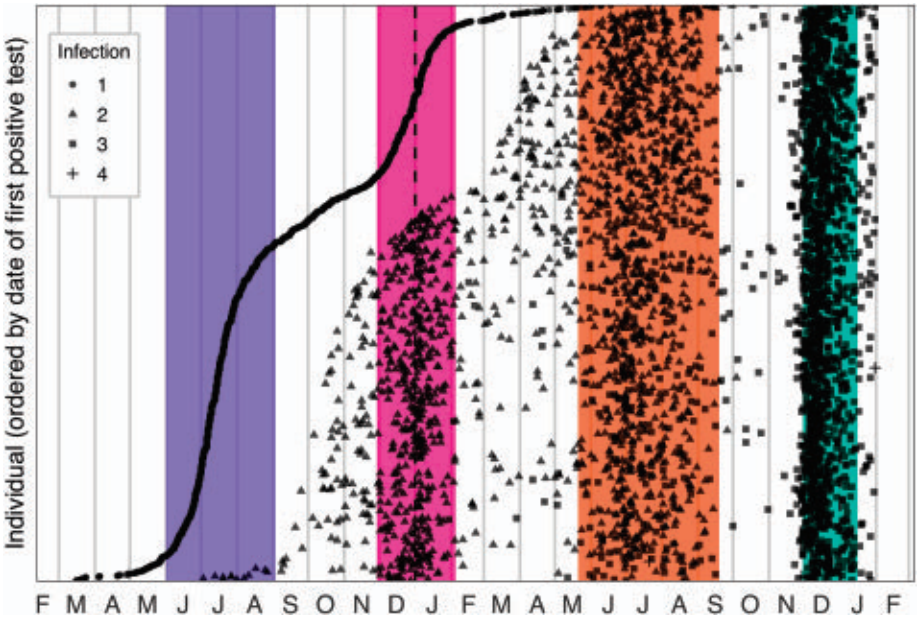


Fig. 3. Timing of infections for individuals with multiple suspected reinfections. Circles represent the first positive test of the first detected infection; triangles represent the first positive test of the suspected second infection; squares represent the first positive test of the suspected third infection; and crosses represent the first positive test of the suspected fourth infection. Colored bands represent wave periods, defined as the period for which the 7-day moving average of cases was at least 15% of the corresponding wave peak (purple indicates wave 1; pink, wave 2; orange, wave 3; and turquoise, wave 4).

relative hazard ratio for the period of Omicron emergence (1 November 2021 to the start of the fourth wave) versus wave 1 was 1.75 (95% CI: 1.48 to 2.10), and for wave 4 versus wave 1, it was 1.70 (95% CI: 1.44 to 2.04).

Discussion and limitations

Our analyses suggest that the cumulative number of reinfections observed through the end of wave 3 was consistent with the null model of no change in reinfection risk through time. Furthermore, our findings suggest that the relative hazard of reinfection versus primary infection decreased with each subsequent wave of infections through September 2021, as would be expected if the risk of primary infection increased without a corresponding increase in reinfection risk. Thus, our analyses show no population-level evidence of immune escape associated with emergence of the Beta or Delta variants. By contrast, in November 2021, the number of daily new reinfections spiked and exceeded the 95% projection interval from the null model, and this was accompanied by a notable increase in the hazard ratio for reinfection versus primary infection. The timing of these changes strongly suggests that they were driven by the emergence of the Omicron variant. This finding has now been supported by analyses of reinfection with Omicron in the United Kingdom and Qatar (*14*–*16*).

Differences in the time-varying force of infection, original and subsequent circulating lineages, testing strategies, and vaccine coverage limit the usefulness of direct comparisons of rates of reinfections across countries or studies. However, pre-Omicron reinfection does appear to be relatively uncommon. The PCR-confirmed reinfection rate ranged from 0 to 1.1% across 11 studies included in a systematic review (*17*). Although none of the studies included in the systematic review reported an increasing risk of reinfection over time, the duration of follow-up was less than a year and most studies were completed before the identification of VOCs. Furthermore, all studies predated the emergence of Omicron. Our findings for the period before the emergence of Omicron are consistent with results from the PHIRST-C community cohort study conducted in two locations in South Africa, which found that infection before the second wave provided 84% protection against reinfection during the second (Beta) wave (*18*), which is comparable to estimates of the level of protection against reinfection for wild-type virus from the SIREN study in the United Kingdom (*19*).

A preliminary analysis of reinfection trends in the United Kingdom suggested that the Delta variant may have a higher risk of reinfection compared with the Alpha variant (*20*); however, this analysis did not take into account the temporal trend in the population

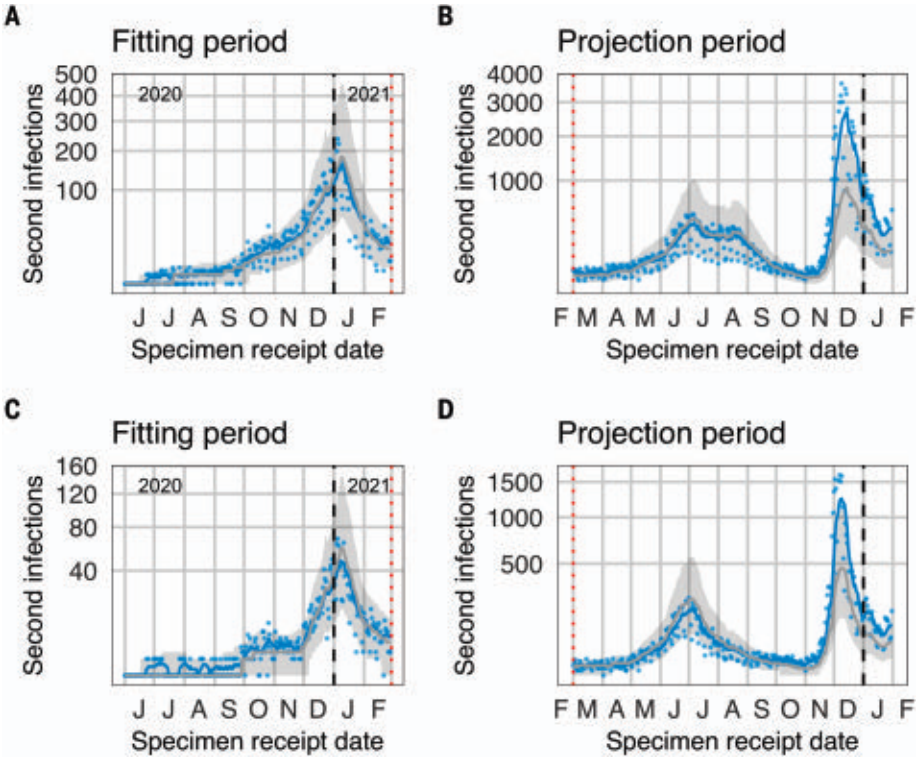


Fig. 4. Observed and expected temporal trends in reinfection numbers. Blue lines (points) represent the 7-day moving average (daily values) of suspected reinfections. Gray lines (bands) represent median predictions (95% projection intervals) from the null model. (**A** and **C**) Null model fit to the data on suspected reinfections through 28 February 2021. (**B** and **D**) Comparison of data with projections from the null model over the projection period. The divergence observed in reinfections from the projection interval in November suggests immune escape. (A) and (B) are South African national data and (C) and (D) are from Gauteng Province only.

at risk for reinfection, which may have biased the findings. Because the Alpha variant never dominated transmission in South Africa, we are unable to analyze the relative risk of reinfection for the Alpha and Delta variants in this context; however, data from Qatar suggest that protection provided by prior infection is similar for Alpha and Delta (*14*).

Our findings regarding the Beta and Delta variants are somewhat at odds with in vitro neutralization studies. Both the Beta and Delta variants are associated with decreased neutralization by some anti–receptor-binding domain and anti–N-terminal domain monoclonal antibodies, although both Beta and Delta remain responsive to at least one anti–receptor-binding domain (*8*, *9*, *21*). In addition, Beta and Delta are relatively poorly neutralized by convalescent sera obtained from unvaccinated individuals infected with non–VOC virus (*7*–*9*, *21*). Finally, sera obtained from individuals after both one and two doses of the BNT162b2 (Pfizer) or ChAdOx1 (AstraZeneca) vaccines displayed lower neutralization of the Beta and Delta variants compared with non-VOCs and the Alpha VOC (*9*). Although this does not have direct bearing on reinfection risk, it is an important consideration for evaluating immune escape more broadly. Non-neutralizing antibodies and T-cell responses could explain the apparent disjuncture between our findings and the in vitro immune evasion demonstrated by both Beta and Delta.

Strengths of this study

Our study has three major strengths. First, we analyzed a large routine national dataset comprising all confirmed cases in the country, allowing a comprehensive analysis of suspected reinfections in the country. Second, we found consistent results using two different analytical methods, both of which accounted for the changing force of infection and increasing numbers of individuals at risk for reinfection. Third, our real-time routine monitoring was sufficient to detect a population-level signal of immune evasion during the initial period of emergence of the Omicron variant in South Africa before results from laboratory-based neutralization tests, providing timely information of importance to global public health planning.

Limitations of this study

The primary limitation of this study is that changes in testing practices, health-seeking behavior, or access to care have not been directly accounted for in these analyses. Estimates based on serological data from blood donors suggest substantial geographic variability in detection rates (*22*), which may contribute to the observed differences in reinfection patterns by province (fig. S1). Detection rates likely also vary through time and by other factors affecting access to testing, which may include occupation, age, and socioeconomic status. In particular, rapid antigen tests, which were introduced in South Africa in late 2020, may be underreported despite mandatory reporting requirements. Although we have incorporated adjustments that account for late reporting of antigen tests, if underreporting of antigen tests were substantial and time varying, then it could still influence our findings. However, comparing temporal trends in infection risk among those eligible for reinfection with the rest of the population, as in approach 2, mitigates against potential failure to detect a substantial increase in risk.

Civil unrest during July 2021 severely disrupted testing in Gauteng and KwaZulu-Natal, the two most populous provinces in the country. Case data are unreliable during the period of unrest and a key assumption of our models, that the force of infection is proportional to the number of positive tests, was violated during this period, resulting in increased misclassification of individuals regarding their status as to whether they were at risk of primary or reinfection. The effect of this misclassification on the signal of immune escape during the period of Omicron's emergence would likely be small and would be expected to bias subsequent reinfection hazard estimates downward.

The purpose of our analysis is to detect changes in the relative reinfection risk through time, rather than to precisely estimate what the reinfection risk is at any particular point in time. Although issues related to underdetection of both primary infections and reinfections are likely to affect the projection intervals against which we compare observed reinfections, we believe that our assessment of changes in the reinfection hazard are fairly robust to these detection issues. In effect, approach 1 follows an open cohort of individuals who have had a first detected infection. Through time, this may include an increasing number of individuals whose first true infection was missed and whose first diagnosed infection was in fact a reinfection. These individuals would presumably be at a reduced risk of acquiring a new infection relative to those whose first detected infection was their first true infection. Two other factors would bias the results in the same direction: undetected reinfections in the cohort of individuals having had a first detected infection and deaths within this cohort, which are not accounted for because of not having a mortality line list

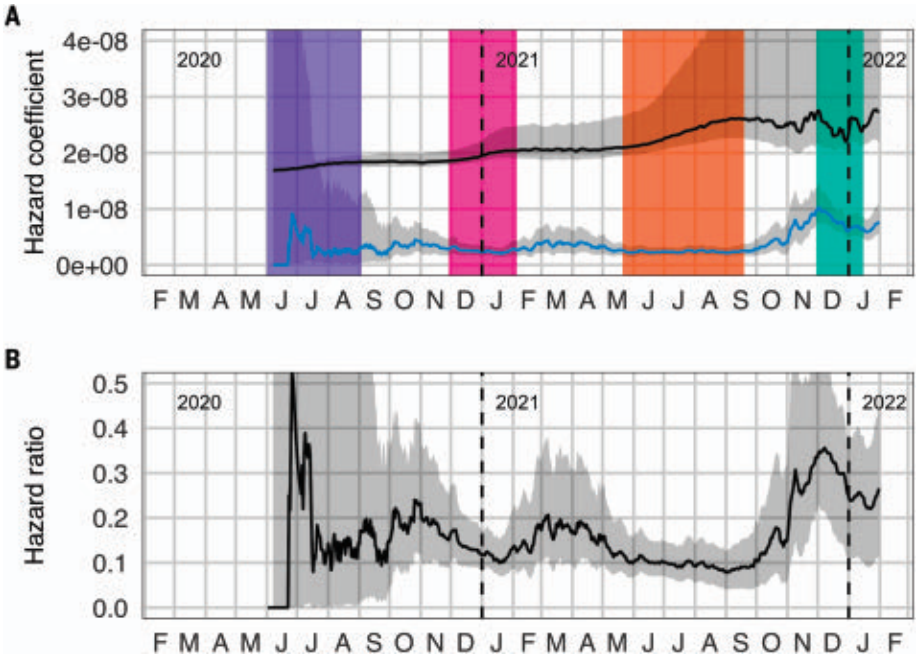


Fig. 5. Estimates of infection and reinfection hazards. (**A**) Estimated time-varying hazard coefficients for primary infection (black) and second infections (blue). Colored bands represent wave periods, defined as the period for which the 7-day moving average of cases was at least 15% of the corresponding wave peak (purple indicates wave 1; pink, wave 2; orange, wave 3; and turquoise, wave 4). (**B**) Ratio of the estimated hazard for reinfections to the estimated hazard for primary infections.

that can be linked to the positive test data. All three factors artificially inflate the estimated denominator of individuals at risk for a second detected infection, thereby reducing the apparent reinfection risk. These factors may explain the slightly lower observed than projected number of reinfections throughout the Delta wave, but did not have a substantial enough effect to prevent detection of the increased reinfection risk associated with the Omicron variant.

The other main limitation of this study is that reinfections were not confirmed by sequencing or by requiring a negative test between putative infections. Nevertheless, the 90-day window between consecutive positive tests reduces the possibility that suspected reinfections were predominantly the result of prolonged viral shedding. Furthermore, because of data limitations, we were unable to determine whether symptoms and severity in primary episodes correlate with protection against subsequent reinfection.

Finally, whereas vaccination may increase protection in previously infected individuals (*23*–*26*), vaccination coverage in South Africa was very low during much of the study period, with only 22.5% of the population fully vaccinated by 30 November 2021 (*27*). Nevertheless, increasing vaccination uptake may reduce the risks of both primary infection and reinfection. The vaccination status of individuals with suspected reinfections identified in this study was unknown. Application of our approach to other locations with higher vaccine coverage would require a more nuanced consideration of the potential effect of vaccination. Further areas for future methodological development include accounting for potential waning of natural and vaccine-derived immunity, as well as methods to track changes in the risk of multiple (three or more) infections.

Given the limitations outlined above, estimates of the extent of immune evasion based on our approach, which aims to detect changing trends rather than make precise estimates, should be treated with caution.

Conclusion

We found evidence of a substantial increase in the risk of reinfection with SARS-CoV-2 that was temporally consistent with the timing of the emergence of the Omicron variant in South Africa, suggesting that Omicron's selection advantage was at least partially driven by an increased ability to infect previously infected individuals.

By contrast, we found no evidence that reinfection risk increased as a result of the emergence of Beta or Delta variants, suggesting that the selective advantage that allowed these variants to spread derived primarily from increased transmissibility rather than from immune evasion. The discrepancy between the population-level evidence presented here and expectations based on laboratory-based neutralization assays for Beta and Delta highlights the need to identify better correlates of immunity for assessing immune escape in vitro.

Immune evasion from prior infection has important implications for public health globally. As new variants emerge, methods to quantify the extent of immune evasion for both natural and vaccine-derived immunity, as well as changes in transmissibility and disease severity, will be urgent priorities to inform facility readiness planning and other public health operations.

Methods

Data sources

Data analyzed in this study came from two sources maintained by the NICD: the outbreak-response component of the Notifiable Medical Conditions Surveillance System (NMC-SS) deduplicated case list and the line list of repeated SARS-CoV-2 tests. All positive tests conducted in South Africa appear in the combined dataset regardless of the reason for testing or type of test (PCR or antigen detection) and include the large number of positive tests that were retrospectively added to the dataset on 23 November 2021 (*28*). Of the 18,585 cases reported on 23 November, 93% had specimen receipt dates before 1 November 2021, and 6% had specimen receipt dates on or after 21 November 2021.

A combination of deterministic (national identity number, name, and date of birth) and probabilistic linkage methods were used to identify repeated tests conducted on the same person. In addition, provincial COVID-19 contact-tracing teams identified and reported repeated SARS-CoV-2–positive tests to the NICD, whether detected by PCR or antigen tests. The unique COVID-19 case identifier that links all tests from the same person was used to merge the two datasets. Irreversibly hashed case IDs were generated for each individual in the merged dataset.

Primary infections and suspected repeat infections were identified using the merged dataset. Repeated case IDs in the line list were identified and used to calculate the time between consecutive positive tests for each individual using specimen receipt dates. If the time between sequential positive tests was at least 90 days, the more recent positive test was considered to indicate a suspected new infection. We present a descriptive analysis of suspected third and fourth infections, although only suspected second infections were considered in the analyses of temporal trends. Incidence time series for primary infections and reinfections were calculated by specimen receipt date of the first positive test associated with the infection, and total observed incidence was calculated as the sum of first infections and reinfections. The specimen receipt date was chosen as the reference point for analysis because it is

complete within the dataset; however, problems have been identified with the accuracy of specimen receipt dates for tests associated with substantially delayed reporting from some laboratories. For these tests, which had equivalent entries for specimen receipt date and specimen report date that were >7 days after the sample collection date, the specimen receipt date was adjusted to be 1 day after the sample collection date, reflecting the median delay across all tests.

All analyses were conducted in the R statistical programming language [version 4.0.5 (2021-03-31)].

Timing of reinfections

We calculated the time between successive infections as the number of days between the last positive test associated with an individual's first or second identified infection (i.e., within 89 days of a previous positive test, if any) and the first positive test associated with their suspected subsequent infection (i.e., at least 90 days after the most recent positive test). We analyzed the distribution of these times for all second and third infections and for the subset of second and third infections occurring since 1 November 2021.

Statistical analysis of reinfection trends

We analyzed the NICD national SARS-CoV-2 routine surveillance data to evaluate whether reinfection risk has changed since the emergence of VOCs in South Africa. We evaluated the daily numbers of suspected reinfections using two approaches. First, we constructed a simple null model based on the assumption that the reinfection hazard experienced by previously diagnosed individuals is proportional to the incidence of detected infections and then fit this model to the pattern of suspected second infections observed through 28 February 2021. The null model assumes no change in the reinfection hazard coefficient through time. We then compared observed reinfections after the fitting period with expected reinfections under projections from the null model. Second, we evaluated whether there has been a change in the relative hazard of reinfection versus primary infection to distinguish between increased overall transmissibility of the variants and any additional risk of reinfection due to potential immune escape. To do this, we calculated a hazard coefficient at each time point for primary and second infections and compared their relative values through time.

Approach 1: Catalytic model assuming a constant reinfection hazard coefficient

Model description

For a case testing positive on day t (by specimen receipt date), we assumed that the reinfection hazard is 0 for each day from $t+1$ to $t+89$ and $\lambda\hat{I}_\tau$ for each day $\tau \geq t+90$, where $\hat{I}_\tau$ is the 7-day moving average of the detected case incidence (first infections and reinfections) for day τ. The probability of a case testing positive on day t having a diagnosed reinfection by day x is thus $p(t,x) = 1 - e^{-\sum_{i=t+90}^{i=x} \lambda\hat{I}_i}$, and the expected number of cases testing positive on day t that have had a diagnosed reinfection by day x is $I_t^1 p(t,x)$, where I_t^1 is the detected case incidence (putative first infections only) for day t. Thus, the expected cumulative number of reinfections by day x is $Y_x = \sum_{t=0}^{t=x} I_t^1 p(t,x)$, and the expected daily incidence of reinfections on day x is $D_x = Y_x - Y_{x-1}$.

Model fitting

The model was fitted to observed reinfection incidence through 28 February 2021 assuming that data are negative binomially distributed with mean D_x. The reinfection hazard coefficient (λ) and the inverse of the negative binomial dispersion parameter (κ) were fitted to the data using an MCMC estimation procedure implemented in the R statistical programming language. We ran four MCMC chains with random starting values for a total of 10,000 iterations per chain, discarding the first 1000 iterations (burn-in). Convergence was assessed using the Gelman-Rubin diagnostic (*29*).

Model-based projection

We used 1500 samples from the joint posterior distribution of fitted model parameters to simulate possible reinfection time series under the null model, generating 100 stochastic realizations per parameter set. We then calculated projection intervals as the middle 95% of daily reinfection numbers across these simulations. We applied this approach at the national and provincial levels.

Approach 2: Estimation of time-varying infection and reinfection hazards

We estimated the time-varying empirical hazard of infection as the daily incidence per susceptible individual. This approach requires reconstruction of the number of susceptible individuals through time. We distinguish between three "susceptible" groups: naive individuals who have not yet been infected (S_1), previously infected individuals who had undetected infections at least 90 days ago and have not yet had a second infection (S_2^u), and previously infected individuals who had a prior positive test at least 90 days ago and have not yet had a second infection (S_2). We estimated the numbers of individuals in each of these categories on day t as follows:

$$S_1(t) = N - \sum_{i=0}^{i=t} P(t)$$

where N is the total population size and $P(t) = P_{\text{obs}}(t)/p_{\text{obs}}$ is the total number of primary infections on day t, of which $P_{\text{obs}}(t)$ were observed and $P_{\text{missed}}(t) = P(t) - P_{\text{obs}}(t)$ were missed.

$$S_2^u(t) = \sum_{i=0}^{i=t-89} P_{missed}(i) - \sum_{i=0}^{i=t-1} U(i)$$

where $U(t) = h_2(t)S_2^u(t)$ is the number of new infections among individuals whose first infection was missed. These individuals are assumed to experience the same infection hazard as individuals whose primary infection was diagnosed and have not yet been reinfected, estimated as $h_2(t) = \frac{\hat{X}_t/p_{obs_2}}{S_2(t)}$. Because individuals are not eligible for reinfection until at least 90 days after their primary infection, we set $U(t) = h_2(t) = 0$ when $t < 90$.

$$S_2(t) = \sum_{i=0}^{i=t-89} P_{obs}(i) - \sum_{i=0}^{i=t} \frac{X_i}{p_{obs_2}}$$

where p_{obs_2} is the probability of detection for individuals who have had a previously identified infection, and X_i is the number of individuals with a second detected infection on day i. Only the possibility of second infections are accounted for in the model, which was developed to monitor reinfection risk against a background in which reinfections were rare.

This setup allows recursive calculation of $U(t)$ and therefore $U_{obs}(t) = U(t)p_{obs_u}$, where p_{obs_u} is the probability of a second infection being observed in an individual whose first infection was missed, and $P_{\text{obs}}(t) = C_t - U_{\text{obs}}(t)$, where C_t is the number of individuals with their first positive test on day t (i.e., detected cases). The daily hazard of infection for previously uninfected individuals is then estimated as $h_1(t) = \frac{\hat{P}_t}{S_1(t)}$.

If we assume that the hazard of infection is proportional to the 7-day moving average of infection incidence ($\hat{Y}_t = \hat{P}(t) + \hat{U}(t) + \hat{X}_t/p_{obs_2}$), then we can then examine the infectiousness of the virus through time as $\lambda_1(t) = h_1(t)/\hat{Y}_t$ and $\lambda_2(t) = h_2(t)/\hat{Y}_t$. We constructed uncertainty intervals around $\lambda_1(t)$, $\lambda_2(t)$, and their ratio, taking into account both measurement noise and uncertainty in the observation parameters (see the supplementary materials for details).

We also used this approach to construct a dataset with the daily numbers of individuals eligible to have a primary infection [$S_1(t)$] or suspected second infection [$S_2(t)$] by wave. Wave periods were defined as the time surrounding the wave peak for which the 7-day moving average of case numbers was >15% of the wave peak. We then analyzed these data

using a generalized linear mixed model to estimate the relative hazard of infection in the population eligible for suspected second infection compared with the hazard in the population not eligible for suspected second infection. For this analysis, we assume $p_{\text{obs}} = 0.1$ and $p_{obs_2} = 0.5$, which falls within the plausible range of observation probabilities (fig. S8).

Our primary regression model was a Poisson model with a log link function, groupinc = Poisson(μ):

log(μ) ~ group*wave + offset[log(groupsize)] + (day)

The outcome variable (groupinc) was the reconstructed daily number of observed infections in the two groups, $P_{\text{obs}}(t)$ and X_t. Our main interest for this analysis was in whether the relative hazard was higher in the second wave, third wave, prewave period in which Omicron emerged, and/or the fourth wave relative to during the first wave, thus potentially indicating immune evasion. This effect is measured by the interaction term between group and wave. The offset term is used to ensure that the estimated coefficients can be appropriately interpreted as per capita rates. We used day as a proxy for force of infection and reporting patterns and examined models where day was represented as a random effect (to reflect that observed days can be thought of as samples from a theoretical population) and as a fixed effect (to better match the Poisson assumptions). Because focal estimates from the two models were indistinguishable, we present only the results based on the random effect assumption.

REFERENCES AND NOTES

1. H. Tegally *et al*., Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature* **592**, 438–443 (2021). doi: 10.1038/s41586-021-03402-9; pmid: 33690265
2. Y. Shu, J. McCauley, GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro Surveill.* **22**, 30494 (2017). doi: 10.2807/1560-7917.ES.2017.22.13.30494; pmid: 28382917
3. R. Viana *et al*., Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. *Nature* 10.1038/s41586-022-04411-y (2022). doi: 10.1038/s41586-022-04411-y; pmid: 35042229
4. National Institute for Communicable Diseases, "The daily COVID-19 effective reproductive number (R) in South Africa: Week 47 2021" (National Institute for Communicable Diseases, 2021); https://www.nicd.ac.za/wp-content/uploads/2021/11/COVID-19-Effective-Reproductive-Number-in-South-Africa-week-47.pdf.
5. National Institute for Communicable Diseases, "The daily COVID-19 effective reproductive number (R) in South Africa: Week 51 2021" (National Institute for Communicable Diseases, 2021); https://www.nicd.ac.za/wp-content/uploads/2021/12/COVID-19-Effective-Reproductive-Number-in-South-Africa-week-51.pdf.
6. C. A. B. Pearson, S. P. Silal, M. W. Z. Li, J. Dushoff, B. M. Bolker, S. Abbott, C. van Schalkwyk, N. G. Davies, R. C. Barnard, W. J. Edmunds, J. Bingham, G. Meyer-Rath, L. Jamieson, A. Glass, N. Wolter, N. Govender, W. S. Stevens, L. Scott, K. Mlisana, H. Moultrie, J. R. C. Pulliam, Bounding the levels of transmissibility & immune evasion of the Omicron variant in South Africa. medRxiv 2021.12.19.21268038 [Preprint] (2021); doi: 10.1101/2021.12.19.21268038
7. S. Cele *et al*., Escape of SARS-CoV-2 501Y.V2 from neutralization by convalescent plasma. *Nature* **593**, 142–146 (2021). doi: 10.1038/s41586-021-03471-w; pmid: 33780970
8. C. K. Wibmer *et al*., SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. *Nat. Med.* **27**, 622–625 (2021). doi: 10.1038/s41591-021-01285-x; pmid: 33654292
9. D. Planas *et al*., Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. *Nature* **596**, 276–280 (2021). doi: 10.1038/s41586-021-03777-9; pmid: 34237773
10. C. Liu *et al*., Reduced neutralization of SARS-CoV-2 B.1.617 by vaccine and convalescent serum. *Cell* **184**, 4220–4236.e13 (2021). doi: 10.1016/j.cell.2021.06.020; pmid: 34242578
11. A. Rössler, L. Riepler, D. Bante, D. von Laer, J. Kimpel, SARS-CoV-2 Omicron variant neutralization in serum from vaccinated and convalescent persons. *N. Engl. J. Med.* **386**, 698–700 (2022). doi: 10.1056/NEJMc2119236; pmid: 35021005
12. D. Planas *et al*., Considerable escape of SARS-CoV-2 Omicron to antibody neutralization. *Nature* **602**, 671–675 (2022). doi: 10.1038/s41586-021-04389-z; pmid: 35016199
13. J. R. C. Pulliam, C. van Schalkwyk, N. Govender, A. von Gottberg, C. Cohen, M. J. Groome, J. Dushoff, K. Mlisana, H. Moultrie, SARS-CoV-2 reinfection trends in South Africa: Analysis of routine surveillance data. medRxiv [Preprint] 2021.11.11.21266068v1 (2021); doi: 10.1101/2021.11.11.21266068v1
14. H. N. Altarawneh *et al*., Protection against the Omicron variant from previous SARS-CoV-2 infection. *N. Engl. J. Med.* 10.1056/NEJMc2200133 (2022). doi: 10.1056/NEJMc2200133; pmid: 35139269
15. N. Ferguson, A. Ghani, A. Cori, A. Hogan, W. Hinsley, E. Volz, Imperial College COVID-19 Response Team, "Report 49: Growth, population distribution and immune escape of Omicron in England" (WHO Collaborating Centre for Infectious Disease Modelling, MRC Centre for Global Infectious Disease Analysis, Jameel Institute, Imperial College London, 2021); https://www.imperial.ac.uk/media/imperial-college/medicine/mrc-gida/2021-12-16-COVID19-Report-49.pdf.
16. UK Health Security Agency, "SARS-CoV-2 variants of concern and variants under investigation in England: Technical briefing 34" (UK Health Security Agency, 2022); https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1050236/technical-briefing-34-14-january-2022.pdf.
17. E. O Murchu *et al*., Quantifying the risk of SARS-CoV-2 reinfection over time. *Rev. Med. Virol.* **32**, e2260 (2022). doi: 10.1002/rmv.2260; pmid: 34043841
18. C. Cohen, J. Kleynhans, A. von Gottberg, M. L. McMorrow, N. Wolter, J. N. Bhiman, J. Moyes, M. du Plessis, M. Carrim, A. Buys, N. A. Martinson, K. Kahn, S. Tollman, L. Lebina, F. Wafawanaka, J. du Toit, F. X. Gómez-Olivé, F. S. Dawood, T. Mkhencele, K. Sun, C. Viboud, for the PHIRST group, S. Tempia, SARS-CoV-2 incidence, transmission and reinfection in a rural and an urban setting: Results of the PHIRST-C cohort study, South Africa, 2020-2021. medRxiv [Preprint] 10.1101/2021.07.20.21260855 (2021); doi: 10.1101/2021.07.20.21260855
19. V. J. Hall *et al*., SARS-CoV-2 infection rates of antibody-positive compared with antibody-negative health-care workers in England: A large, multicentre, prospective cohort study (SIREN). *Lancet* **397**, 1459–1469 (2021). doi: 10.1016/S0140-6736(21)00675-9; pmid: 33844963
20. Public Health England, "SARS-CoV-2 variants of concern and variants under investigation in England" (Public Health England, 2021); https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1005517/Technical_Briefing_19.pdf.
21. P. Wang *et al*., Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. *Nature* **593**, 130–135 (2021). doi: 10.1038/s41586-021-03398-2; pmid: 33684923
22. M. Vermeulen, L. Mhlanga, W. Sykes, C. Coleman, N. Pietersen, R. Cable, R. Swanevelder, T. N. Glatt, E. Grebe, A. Welte, K. van den Berg, Prevalence of anti-SARS-CoV-2 antibodies among blood donors in South Africa during the period January-May 2021. Research Square [Preprint] (2021); https://www.researchsquare.com/article/rs-690372/v2. doi: 10.21203/rs.3.rs-690372/v2
23. L. Stamatatos *et al*., mRNA vaccination boosts cross-variant neutralizing antibodies elicited by SARS-CoV-2 infection. *Science* **372**, 1413–1418 (2021). doi: 10.1126/science.abg9175; pmid: 33766944
24. F. Krammer *et al*., Antibody responses in seropositive persons after a single dose of SARS-CoV-2 mRNA vaccine. *N. Engl. J. Med.* **384**, 1372–1374 (2021). doi: 10.1056/NEJMc2101667; pmid: 33691060
25. S. Saadat *et al*., Binding and neutralization antibody titers after a single vaccine dose in health care workers previously infected With SARS-CoV-2. *JAMA* **325**, 1467–1469 (2021). doi: 10.1001/jama.2021.3341; pmid: 33646292
26. Y. Lustig *et al*., Neutralizing response against variants after SARS-CoV-2 infection and one dose of BNT162b2. *N. Engl. J. Med.* **384**, 2453–2454 (2021). doi: 10.1056/NEJMc2104036; pmid: 33826815
27. Department of Health – South Africa, "Update on Covid-19 (09th March 2021)" (Department of Health – South Africa, 2021); https://sacoronavirus.co.za/2021/03/09/update-on-covid-19-09th-march-2021/).
28. South African Government News Agency, "Health Department adds antigen tests to country's official COVID-19 stats" (South African Government News Agency, 2021); https://www.sanews.gov.za/south-africa/health-department-adds-antigen-tests-countrys-official-covid-19-stats).
29. A. Gelman, D. B. Rubin, Inference from iterative simulation using multiple sequences. *Stat. Sci.* **7**, 457–472 (1992). doi: 10.1214/ss/1177011136
30. J. R. C. Pulliam, C. van Schalkwyk, N. Govender, A. von Gottberg, C. Cohen, M. J. Groome, J. Dushoff, K. Mlisana, H. Moultrie, Data for "Increased risk of SARS-CoV-2 reinfection associated with emergence of Omicron in South Africa," Zenodo (2022); doi: 10.5281/zenodo.6108448
31. J. R. C. Pulliam, C. van Schalkwyk, N. Govender, A. von Gottberg, C. Cohen, M. J. Groome, J. Dushoff, K. Mlisana, H. Moultrie, Code for "Increased risk of SARS-CoV-2 reinfection associated with emergence of Omicron in South Africa," Zenodo (2022); doi: 10.5281/zenodo.5807591

ACKNOWLEDGMENTS

We thank the members of the NICD Epidemiology and Information Technology teams who curated, cleaned, and prepared the data used in this analysis (epidemiology team: A. Moipone Shonhiwa, G. Ntshoe, J. Ebonwu, L. Motsuku, L. Shuping, M. Muchengeti, J. Kleynhans, G. Hunt, V. Odhiambo Olago, H. Ismail, N. Govender, A. Mathews, V. Essel, V. Msimang, T. Kufa-Chakezha, N. Villyen Motaze, N. Mayet, T. Mmaborwa Matjokotja, M. Neti, T. Arendse, T. Lamola, I. Matiea, D. Muganhiri, B. Ndlovu, K. Ravhuhali, E. Ramutshila, S. Mhlanga, A. Mzoneli, N. Naran, T. Whitbread, M. Moeti, C. Iwu, E. Mathatha, F. Gavhi, M. Makamu, M. Makhubele, S. Mdleleni, B. Chiger, and J. Kleynhans; information technology team: T. Mukange, T. Bell, L. Darwin, F. McKenna, N. Munava, M. Raza Bano, T. Ngobeni). We also thank C. A. B. Pearson, S. Horn, Y. Jo, B. Lombard, L. S. Villabona-Arenas, and colleagues in the South African COVID-19 Modelling Consortium and the SARS-CoV-2 variants research consortium in South Africa for helpful discussions during the development of this work. In addition, we acknowledge the Network for Genomic Surveillance - South Africa (NGS-SA) led by T. de Oliveira for its role in discovery of the Omicron variant. **Funding:** J.R.C.P. and C.v.S. are supported by the South African Department of Science and Innovation and the National Research Foundation. Any opinion, finding, and conclusion or recommendation expressed in this material is that of the authors, and the NRF does not accept any liability in this regard. This work was also supported by the Wellcome Trust (grant no. 221003/Z/20/Z) in collaboration with the Foreign, Commonwealth and Development Office, United Kingdom. **Author contributions:** Conceptualization: J.R.C.P., C.v.S., J.D., H.M.; Data curation: N.G., K.M., A.v.G., C.C.; Formal analysis: J.R.C.P., C.v.S., J.D.; Investigation: J.R.C.P., C.v.S., A.v.G., C.C., M.J.G., J.D., H.M.; Writing – original draft: J.R.C.P., C.v.S., H.M.; Writing – review & editing: all authors. **Competing interests:** All authors have completed the International Committee of Medical Journal Editors (ICMJE) uniform disclosure form. C.C. and A.v.G. have received funding from Sanofi Pasteur in the past 36 months. J.R.C.P. and K.M. serve on the Ministerial Advisory Committee on COVID-19 of the South African National Department of Health. The authors declare no other relationships or activities that could appear to have influenced the submitted work. Ethical approval: This study has received ethical clearance from University of the Witwatersrand (clearance certificate no. M210752, formerly M160667) and approval under reciprocal review from Stellenbosch University (project ID 19330, ethics reference no. N20/11/074_RECIP_WITS_M160667_COVID-19). **Data availability:** Data and code are available on Zenodo (*30*, *31*). The following data are included in the repository: counts of reinfections and primary infections by province, age group (5-year bands), and sex (M, F, U); daily time series of primary infections and suspected reinfections by specimen receipt date (national); and model output: posterior samples from the MCMC fitting procedure and simulation results. All other data are covered by a nondisclosure agreement and cannot be released by the authors. Requests for additional data must be made in writing to the National Institute

for Communicable Diseases, South Africa. Requests should be directed to H.M. (harrym@nicd.ac.za). 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.abn4947
Supplementary Text
Figs. S1 to S9
Tables S1 and S2
References (*32*, *33*)
MDAR Reproducibility Checklist

Submitted 30 November 2021; accepted 9 March 2022
Published online 15 March 2022
10.1126/science.abn4947

RESEARCH ARTICLES

CLIMATE ECONOMICS

Land-use emissions embodied in international trade

Chaopeng Hong[1,2]*†, Hongyan Zhao[3,4]†, Yue Qin[5], Jennifer A. Burney[6], Julia Pongratz[7,8], Kerstin Hartung[7]‡, Yu Liu[9,10], Frances C. Moore[11], Robert B. Jackson[12], Qiang Zhang[4], Steven J. Davis[2,13]*

International trade separates consumption of goods from related environmental impacts, including greenhouse gas emissions from agriculture and land-use change (together referred to as "land-use emissions"). Through use of new emissions estimates and a multiregional input-output model, we evaluated land-use emissions embodied in global trade from 2004 to 2017. Annually, 27% of land-use emissions and 22% of agricultural land are related to agricultural products ultimately consumed in a different region from where they were produced. Roughly three-quarters of embodied emissions are from land-use change, with the largest transfers from lower-income countries such as Brazil, Indonesia, and Argentina to more industrialized regions such as Europe, the United States, and China. Mitigation of global land-use emissions and sustainable development may thus depend on improving the transparency of supply chains.

Human land use, while producing vast quantities of agricultural and forestry products, has also disrupted ecosystems (*1*), degraded biodiversity (*2*), and added considerable quantities of greenhouse gases (GHGs) to the atmosphere (*3*). However, as a result of international trade, environmental effects have often occurred in different regions from where products are consumed (*4*, *5*). Previous studies have thus quantified energy-related emissions of CO_2 embodied in global trade (i.e., traded goods and services) (*6*, *7*), as well as air pollution (*8*) and land (*9*–*11*) and water resources (*12*, *13*) virtually embodied in trade. However, previous assessments of land-use GHG emissions embodied in trade have focused on specific regions and commodities such as Brazilian cattle and soybeans (*14*) or meat (*15*), with no comprehensive global analysis of emissions from both agriculture and land-use change.

Global analysis of trade-related land-use emissions has been hindered by a lack of sufficiently detailed estimates of country-, region-, and product-specific land-use emissions [including emissions from both agriculture and land-use change (LUC)] (*14*, *16*, *17*). However, a recent study (*3*) has now supplied the necessary data showing that land-use emissions represent ~25% of net anthropogenic GHG emissions in recent years, about half the amount as methane (CH_4) and nitrous oxide (N_2O) emissions from agricultural production (including both crops and livestock) and half as CO_2 emissions from LUC (*3*, *18*–*21*). These land-use emissions are substantial enough to threaten international climate goals even if fossil fuel emissions are drastically reduced (*3*, *22*). Trade-related accounting of global land-use emissions is needed to reveal the international drivers of land-use emissions and to better target and coordinate mitigation efforts.

We present comprehensive estimates of land-use emissions embodied in international trade. Details of our analytic approach are described in the materials and methods (*23*). In summary, we use a multiregional input-output model to attribute land-use emissions associated with the production of various agricultural and forestry products to final consumption in 141 world regions (most individual countries) in the years 2004, 2007, 2011, 2014, and 2017. Trade data are from the Global Trade Analysis Project (GTAP) and production-based estimates of land-use emissions and agricultural products are from Hong *et al.* (*3*), which are in turn based on agricultural emissions from the United Nations Food and Agriculture Organization (FAO) (*24*) and LUC emissions from a spatially explicit bookkeeping model (BLUE) (*16*, *21*). By using the "base case" accounting assumptions of Hong *et al.* (*3*), we report emissions in units of CO_2 equivalent (CO_2-eq) with the 100-year global warming potentials of CH_4 and N_2O and assign land-use

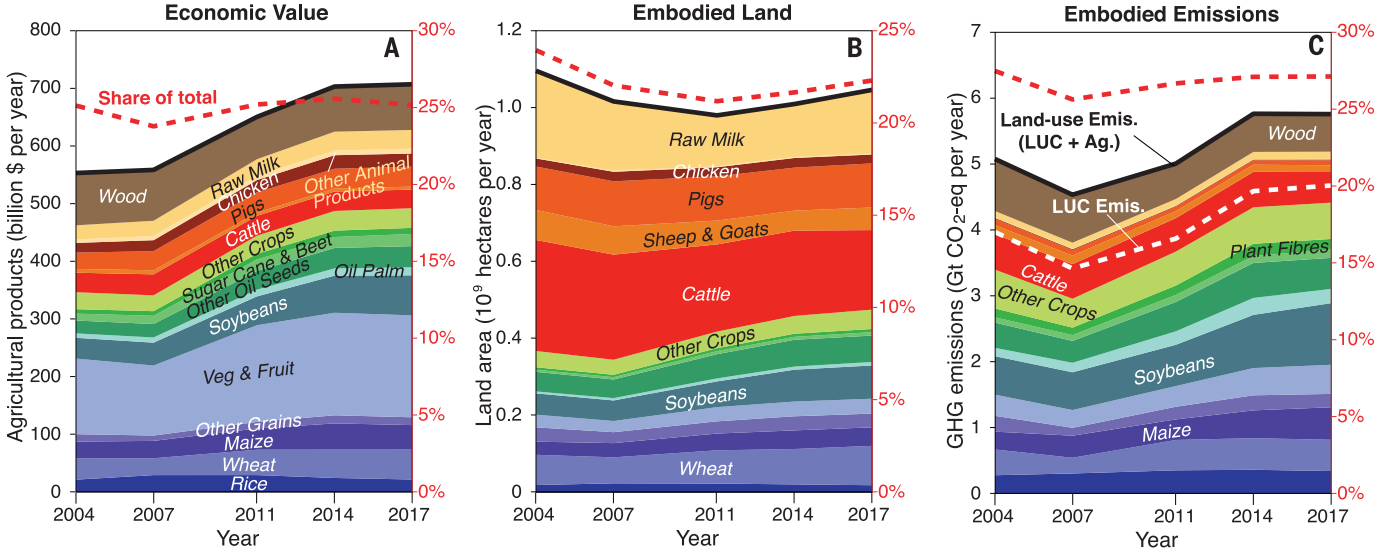


Fig. 1. Agricultural products, land use, and land-use emissions embodied in international trade. (**A**) Global agricultural and forestry products, (**B**) agricultural land use, and (**C**) land-use emissions embodied in trade over 2004 to 2017 by product (sector). In each panel, the global total is indicated by the bold black line. The right *y* axis shows what share of total agricultural value, land use, and emissions are embodied in trade. In (A), the agricultural products traded are in units of constant 2004-2006 billion international dollars. In (C), the subset of land-use change (LUC) emissions is indicated by the dashed white line.

change emissions to the years in which they probably occurred ("legacy" emissions). We exclude carbon uptake related to abandoned agricultural land, as it is not possible to associate such land with a specific product. The input-output approach—widely used for calculating embodied emissions (*6*, *7*)—maps emissions to the region where related goods are ultimately consumed (consumption emissions), even if the products are transshipped through an intermediary region or are intermediate constituents in a multiregional supply chain. Therefore, our trade analysis provides valuable information beyond bilateral trade data, which do not track such re-exports (*12*, *25*). The difference between each region's production emissions and consumption emissions represents the net emissions embodied in trade and is therefore equal to the emissions related to imported goods minus the emissions related to exported goods. Lastly, we analyze the driving factors of country- or region-level changes in embodied emissions from 2004 to 2017 by a structural decomposition analysis. Our analysis is focused on land-use emissions embodied in international trade at the national scale; thus, subnational and smaller-scale details are not well revealed, though such local processes will ultimately be critical to more sustainable global food production and trade.

Land use and emissions embodied in trade

Between 2004 and 2017, we found that roughly 1 billion hectares of agricultural land (both cropland and pasture) were used for traded agricultural products, representing ~22% of agricultural land worldwide (Fig. 1B), a result consistent with previous studies (*11*, *26*). However, a somewhat higher share of global land-use emissions (~27%) are embodied in international trade, ranging between 4.5 and 5.8 billion metric tons (Gt) CO_2-eq per year during the study period (note that carbon uptake from agriculture abandonment is not included in the analysis) (Fig. 1C). Among the traded commodities, cereals (rice, wheat, maize, and other grains) and oil crops (soybeans, oil palm, and other oil seeds) together represent 26 to 35% of the land use of traded products and 45 to 54% of embodied emissions over the study period. Animal products (such as cattle, sheep, pigs, chicken, and raw milk), which generally require more land area per unit produced (*3*, *27*), constitute 55 to 67% of embodied land use but only 14 to 19% of net embodied emissions. By contrast, although vegetables and fruits represent a large share (roughly a quarter) of traded agricultural and forestry products by value (Fig. 1A), their land requirements and emissions are comparatively small (representing <8% of embodied land and emissions).

The dominant global feature of embodied land-use emissions, persistent over time, are large exports of emissions from countries such as Brazil, Indonesia, Argentina, Australia, and Canada to consumers in developed regions such as the US, Europe, and Japan (Fig. 2). However, there were some substantial changes over the study period: although China was an important net exporter of agricultural products in 2004, rapid growth in imports during the study period meant that by 2017 it was an

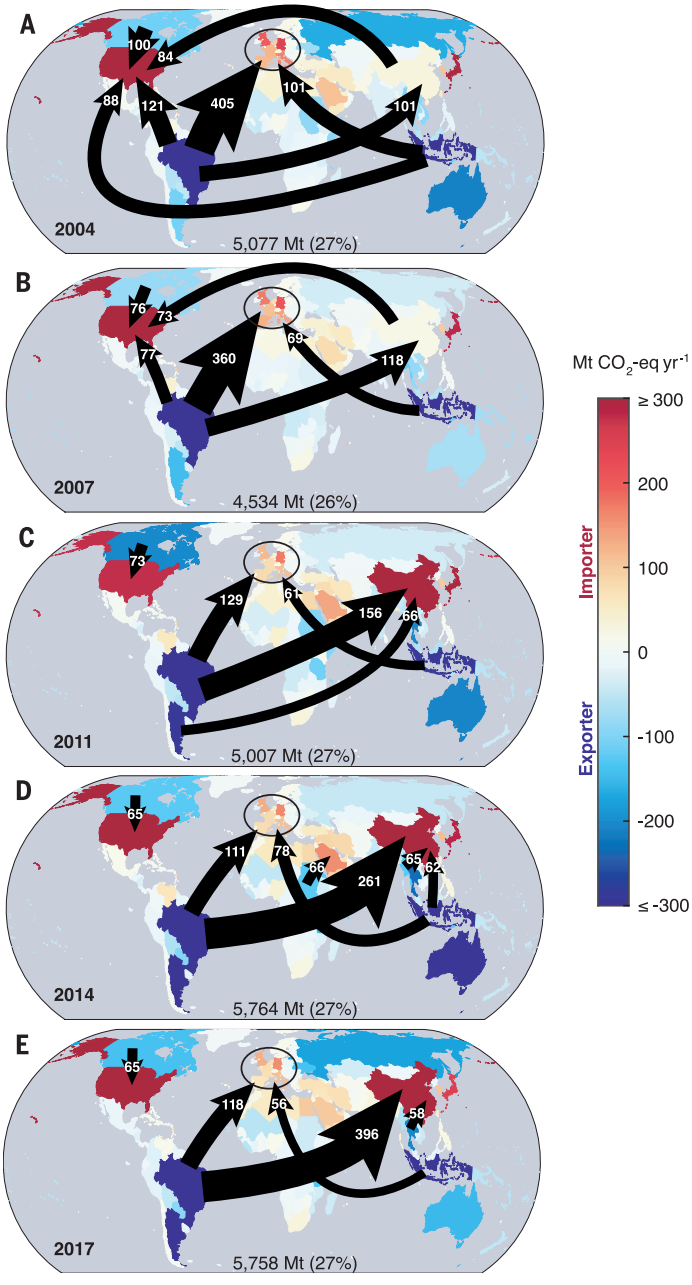


Fig. 2. Global distribution of land-use emissions embodied in trade. (**A** to **E**) Net land-use emissions embodied in trade for each region (shading) and largest net fluxes of embodied emissions between regions (arrows) in 2004 (A), 2007 (B), 2011 (C), 2014 (D), and 2017 (E). Shading indicates the magnitude of net emissions embodied in trade with net exporters blue and net importers red. Fluxes to and from Europe are aggregated to include all 27 member states of the European Union plus the UK. The scale of arrows across the maps is consistent except for arrows with values >200 Mt CO_2-eq $year^{-1}$, which are drawn at the same size.

[1]Institute of Environment and Ecology, Shenzhen International Graduate School, Tsinghua University, Shenzhen, China. [2]Department of Earth System Science, University of California, Irvine, Irvine, CA, USA. [3]School of Environment, Beijing Normal University, Beijing, China. [4]Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing, China. [5]College of Environmental Science and Engineering, Peking University, Beijing, China. [6]School of Global Policy and Strategy, University of California, San Diego, San Diego, CA, USA. [7]Department of Geography, Ludwig-Maximilians-Universität, Munich, Germany. [8]Max Planck Institute for Meteorology, Hamburg, Germany. [9]Institute of Science and Development, Chinese Academy of Sciences, Beijing, China. [10]School of Public Policy and Management, University of Chinese Academy of Sciences, Beijing, China. [11]Department of Environmental Science and Policy, University of California, Davis, Davis, CA, USA. [12]Department of Earth System Science, Woods Institute for the Environment, and Precourt Institute for Energy, Stanford University, Stanford, CA, USA. [13]Department of Civil and Environmental Engineering, University of California, Irvine, Irvine, CA, USA.
***Corresponding author. Email: hcp12@tsinghua.org.cn (C.H.); sjdavis@uci.edu (S.J.D.)**
†These authors contributed equally to this work.
‡Present address: Deutsches Zentrum für Luft- und Raumfahrt, Institut für Physik der Atmosphäre, Oberpfaffenhofen, Germany.

important importer of agricultural products by value, as well as the largest net importer of embodied land use and emissions (Figs. 2 and 3 and figs. S1 and S2). The growth in exports from Brazil to China—one of the most important international trade flows of land-use emissions in 2017—coincided with diminishing Brazilian exports to Europe and the US, one of the most important trade flows in 2004 (Fig. 2). The value of traded products and the land required to produce them (figs. S1 and S2) show similar changes over the decade of study. Figure S1 shows that the US was a net importer of agricultural products by value in 2004 but was a net exporter of agricultural products in 2014. Indonesia, by contrast, was a net exporter of agricultural products by value and embodied emissions but a net importer of embodied land use. Finally, Russia was a net importer of agricultural products by value and embodied land use before 2014 but was a net exporter in 2017, attributable to its changing trade policies (e.g., import bans on certain agricultural products from some Western countries).

We further decompose embodied land-use emissions into those related to land-use change (i.e., CO_2 emissions related to land conversion; LUC) versus agricultural processes (i.e., GHGs emitted directly from the production of crops and livestock; see materials and methods for details) (*23*). Of the total global embodied land-use emissions over the study period, 75 to 81% are related to land-use change (dashed white line in Fig. 1C), particularly conversion to croplands, suggesting that international trade in agricultural commodities (especially crops) is related to substantial land-use change and related CO_2 emissions (*28*–*31*). Indeed, the major trade flows of land-use emissions (e.g., from Brazil and Indonesia) are also dominated by LUC emissions (Fig. 2 and fig. S3). Of all agricultural emissions embodied in trade, roughly two-thirds are related to CH_4 (from sources such as enteric fermentation, rice cultivation, and manure management) and one-third is related to N_2O (e.g., from fertilizers and manure). The pattern of exporters and importers of agricultural emissions is somewhat similar to that of LUC emissions, but there are also differences (figs. S3 and S4) that reflect considerable variation in emissions intensities across regions.

Regional land-use emissions

Brazil was the largest net exporter of land-use emissions in 2017 [917 million tons (Mt) CO_2-eq], followed by Argentina, Indonesia, Thailand, Russia, and Australia (150 to 330 Mt CO_2-eq each; Fig. 3J). The largest net importer of land-use emissions in 2017 was China (814 Mt CO_2-eq), followed by the US, Japan, Germany, the UK, Italy, South Korea, and Saudi Arabia (100 to 370 Mt CO_2-eq each; Fig. 3G). There is substantial but imperfect overlap between the largest exporters and importers of emissions and the largest exporters and importers of land use. For example, the largest net exporter of agricultural land use in 2017 was Australia [200 million hectares (Mha)], followed by Brazil, Mongolia, Argentina, and Canada (24 to 75 Mha each; Fig. 3K); notably, many of these countries are important exporters of animal products. On a per capita basis, net exports of emissions are largest (5 to 13 tons per capita) in regions where agriculture also accounts for a large share of regional economies and exports (e.g., Paraguay, Australia, and Argentina; Fig. 3L). In these cases, more than half of the land-use emissions produced (and in Australia, more than half of agricultural land use) are related to exports (fig. S5, A and B, and fig. S6B). Similarly, net exports account for 40% of all land-use emissions produced in Brazil, and ~30% of Brazil's agricultural land is used to support exports. By contrast, net imports of emissions are greatest (4 to 8 tons per capita) in affluent regions with little arable land (e.g., United Arab Emirates, Hong Kong, Singapore, Luxembourg, Kuwait; Fig. 3I). In many of these regions, >90% of land-use emissions related to final consumption are embodied in imports (figs. S5D and S6C). The land use embodied in imports to Hong Kong, Singapore, and Bahrain

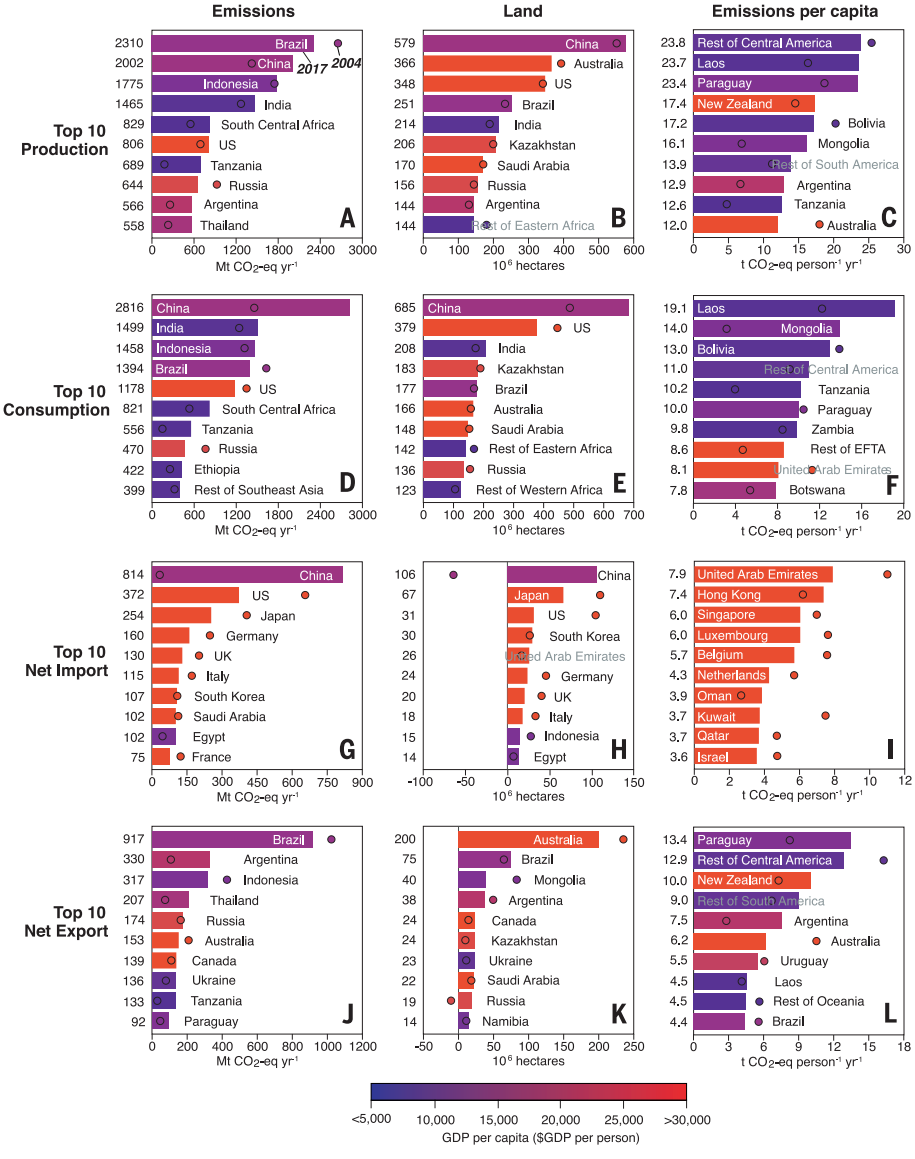


Fig. 3. Top 10 countries or regions by land-use emissions and agricultural land use. (**A** to **L**) Top ten countries/regions in 2017 by production (A) to (C), consumption (D) to (F), net imports (G) to (I) and net exports (J) to (L) of land-use emissions (A), (D), (G), and (J), agricultural land use (B), (E), (H), and (K) and per capita emissions (C), (F), (I), and (L). Data in 2017 are shown as bars shaded by GDP per capita. Data in 2004 are indicated by circles.

is in each case 1 to 2 orders of magnitude as large as the entire area of these regions (fig. S5F).

Exports from Brazil, Indonesia, Argentina, Australia, Thailand, and Tanzania are particularly emissions-intensive (Fig. 4). These agriculturally productive regions have carbon-dense forests (e.g., Brazil and Indonesia) (*32*) and/or produce emissions-intensive products (e.g., sheep and cattle in Australia), and more than three-quarters of their collective land-use emissions are from land-use change. By contrast, the emissions intensity of exports from China, the US, Europe, and Japan is much lower and includes low levels of LUC emissions. Thus, emissions intensity of imports by affluent regions is typically much greater than that of their own exports or domestic production (Fig. 4 and fig. S6D), suggesting that consumption in some affluent regions drives land-use change and emissions in lower-income regions. Indeed, such large differences in emission intensity make the US a net emission importer, although it was a net exporter of agricultural products by value in 2014 (Fig. 2 and fig. S1). China, by contrast, was a net importer of LUC emissions and a net exporter of agricultural emissions from 2004 to 2011. Emission intensity of imports varies much less across regions, though it is noticeably high in China and India (Fig. 4).

The exported land-use emissions from major exporters are usually dominated by a small subset of exported products (Fig. 5). For example, soybeans make up the largest component of emissions exported from Brazil (517 Mt CO_2-eq in 2017) and Argentina (174 Mt CO_2-eq), whereas oil palm makes up the largest component of emissions exported from Indonesia (132 Mt CO_2-eq) and Malaysia (56 Mt CO_2-eq), where agricultural expansion often occurs on peatlands. These products often have a high share of embodied emissions from land-use change to cropland (Fig. 5 and fig. S7). By contrast, animal products such as cattle and sheep represent most of India's exported emissions (~70 Mt CO_2-eq in 2017), with a comparatively large share of the embodied emissions related to agricultural processes [e.g., CH_4 and N_2O from enteric fermentation and manure management (*33*); Fig. 5 and fig. S7]. Rice represents nearly half of the emissions exported by Thailand, with considerable emissions from both land-use change and agricultural processes (CH_4 from rice cultivation). Wheat, cattle, and sheep dominate emissions exported by Australia. Imported emissions are comparatively diversified, though a large share of emissions imported to China are related to oil crops such as soybeans (~600 Mt CO_2-eq in 2017).

Figure S7 shows the 50 largest product-specific trade flows of embodied emissions from 2004 to 2017. This list includes many large trade flows (e.g., Brazil-China, Brazil-Europe, Indonesia-Europe) and major commodity crop and animal products (soybeans, oil palm, cattle, maize), as well as some less recognized sources (e.g., Tanzania-Europe: tobacco, cocoa, coffee, and tea; Cote d'Ivoire-Europe: cocoa; Indonesia/Thailand-China: fiber and rubber). China's soybean imports from Brazil, Argentina, and the US are all among the top 10 trade flows of emissions during the study period. The top flows may thus represent targeted opportunities for efforts to substantially reduce land-use emissions through global supply chains [e.g., expanding the soy moratorium (*34*) to Brazil's Cerrado (*35*) and bolstering and broadening efforts such as the Roundtable on Sustainable Palm Oil (*36*) to include rubber from Southeast Asia and cash crops such as cocoa from Equatorial Africa].

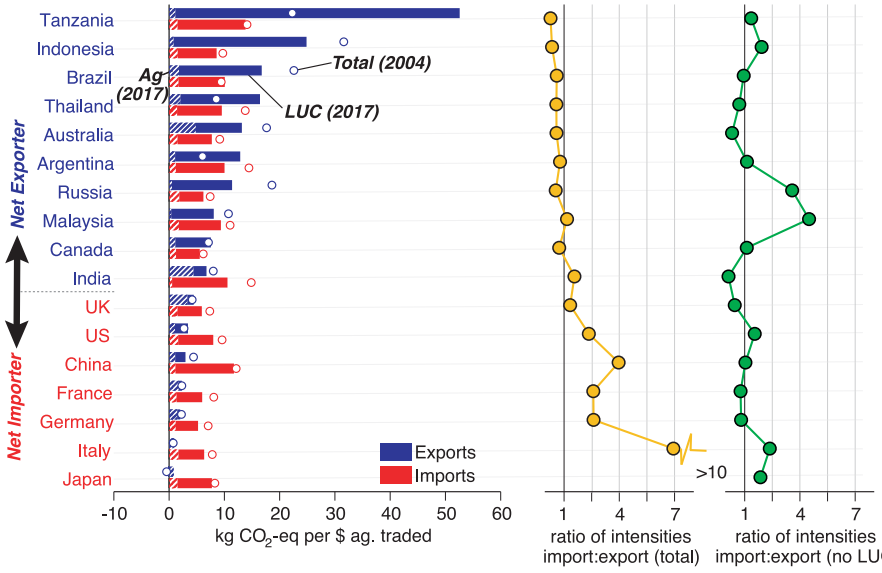


Fig. 4. Land-use emission intensity of exports and imports by major traders. Intensity of total land-use emissions in 2017 is separated into land-use change emissions (LUC; solid bars) and agricultural emissions (Ag; hatched bars). Intensity of total land-use emissions in 2004 is indicated by circles.

Changes and drivers: 2004 to 2017

Land-use emissions embodied in global trade increased by 14% between 2004 and 2017, from 5.1 to 5.8 Gt CO_2-eq per year (Fig. 1C). Meanwhile, the land area supporting traded agricultural products decreased by 5% over the same period, with increases in cropland areas offset by decreases in pasture for livestock (Fig. 1B). Figure 6 decomposes the factors contributing to the changes in annual embodied emissions over the study period, showing that the 0.7-Gt CO_2-eq increase was mostly due to increases in trade volume (+1.6 Gt) partly offset by decreases in emission intensity (−0.9 Gt). In both cases, these changes were dominated by LUC emissions; embodied agricultural emissions changed very little (−0.02 Gt) because steady increases in trade volume (+0.23 Gt) were largely offset by decreases in emission intensity (−0.09 Gt) and changes in trade structure (−0.16 Gt). Although some changes in trade structure—such as the decline in trade of cattle and sheep—reduced embodied agricultural emissions, others acted to increase LUC emissions through expansion of cropland, resulting in an overall small effect on emissions embodied in trade (Figs. 1C and 6A, and fig. S7).

More than half of the increase in annual land-use emissions embodied in trade from 2004 to 2017 were related to exports from Argentina, Thailand, and Tanzania, which together rose by 477 Mt CO_2-eq (+171%), in each case largely driven by increased emission intensity of LUC emissions accompanied with increased volume of agricultural exports (Fig. 6B). Meanwhile, emissions embodied in Brazilian exports decreased by 81 Mt CO_2-eq (−8%), in this case primarily because of decreased emission intensity of LUC emissions related to decreases in deforestation in the Brazilian Amazon between 2004 and 2017 (*34*, *37*, *38*), which offset the growth from increased trade volume (Fig. 6B and fig. S8). By contrast, emissions embodied in imports to China and India increased by 692 Mt CO_2-eq (+206%) and 77 Mt CO_2-eq (+69%), respectively, over the study period, mostly driven by the increasing volume of imports (Fig. 6C and fig. S8). Were it not for the observed changes in emissions intensity (that is, if trade volume and trade structure were the only factors affecting embodied emissions), then annual land-use emissions embodied in trade would have increased by 1.6 Gt CO_2-eq (+31%) from 2004 to 2017,

and Brazil would have the largest increase—rather than decrease—in exported emissions (Fig. 6). Similarly, were it not for LUC emissions (that is, if only agricultural emissions were analyzed), there would have been very little change in emissions embodied in trade (−0.02 Gt; Fig. 6).

Further details of changes in embodied land-use emissions from 2004 to 2017 related to specific region and product combinations are shown in Fig. 5 and figs. S7, S8, S9, and S10. In particular, China's soybean imports and the related embodied emissions tripled from 2004 to 2017, reaching 470 Mt CO_2-eq in 2017—mostly related to increases in exported emissions from Brazil, the US, and Argentina. The changes in soybean imports are a large part of the changing trade flows in Fig. 2; soybeans alone represent 46% of the 287 Mt decrease in Brazil's net exported emissions to Europe from 2004 to 2017, whereas Brazil's net exported emissions to China rose over the same period (Fig. 2 and fig. S7).

Discussion and conclusions

Our results reveal that the share of land-use emissions embodied in international trade in recent years is comparable to the relative share of fossil fuel emissions traded internationally (27 and 21% in 2017, respectively; both calculated by this study, see fig. S11 and materials and methods) (*23*). However, unlike fossil fuel CO_2 emissions, the largest net exporters of land-use emissions are concentrated in agrarian regions of the Southern hemisphere, in which the share of agricultural products exported is often high (e.g., Australia and Argentina). Further, carbon-dense and biodiverse ecosystems are often cleared and managed (e.g., tropical forests in Brazil and peatlands in Indonesia) to support exports of agricultural commodities to more affluent and/or populous regions such as Europe, the US, China, and Japan (Figs. 2 and 4, and fig. S6). Consumption-based land-use emissions are considerably lower than production-based emissions in Oceania, Southeast Asia, and Latin America, whereas the opposite is true in Europe, North America, and East Asia (figs. S12 and S13). Meanwhile, Europe, the US, and Japan are major net importers of both land-use emissions and fossil fuel CO_2 emissions, whereas substantial fossil emissions are embodied in China's exports, offsetting China's imports of land-use emissions (fig. S11). Although embodied land-use emissions are comparable in magnitude to embodied fossil fuel emissions in many regions (fig. S11), consumption-based accounting of land-use emissions is not yet required by any national policies or international agreements (*39*–*41*).

One central question is whether trade of agricultural products has inadvertently increased or decreased global emissions relative to a world with no trade or different trade patterns. Although our methods cannot fully evaluate all trade counterfactuals, a simplistic approach that has been applied in the context of water use (*13*, *42*) compares the emissions actually produced by exporters to the emissions that importers would produce if they were to create the imported products domestically (holding all current emissions intensities constant and assuming continued trade in products importers have never produced). This approach suggests that trade of the top 10 net importers of emissions tends to increase global land-use emissions (their imports are more emissions-intensive than domestic production; fig. S6), whereas imports to Latin American and sub-Saharan African regions tend to reduce global emissions (their production is more emissions-intensive than imports). However, this simple method further suggests that in recent years global trade may have both increased and decreased global land-use emissions (by +1.2 Gt CO_2-eq in 2004 and −0.2 Gt CO_2-eq in 2017, respectively). Given the strong assumptions entailed, these results should be interpreted cautiously, but they indicate that a substantial scale of climate mitigation might be possible through strategic trade adjustments. Policy levers that take comparative environmental advantage into account—such as consumption-based accounting and international price adjustments—may help induce such adjustments.

Regardless of the global effect, however, prevailing patterns of trade and land use suggest that at least some trade flows may be exacerbating global GHG emissions and ecological habitat destruction. This conclusion is supported by a large and diverse literature concerned with "displaced land use" and

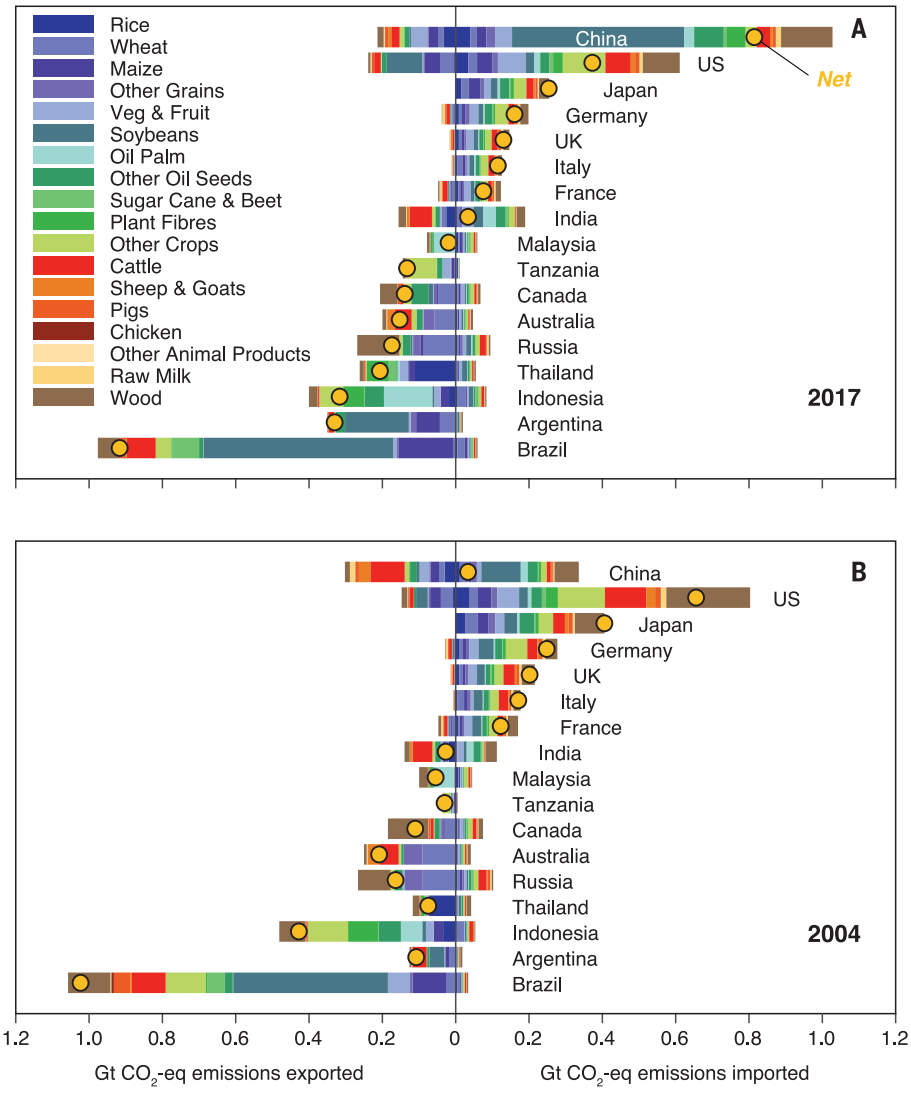


Fig. 5. Balance of land-use emissions embodied in imports and exports of major traders. Trade balance in (**A**) 2017 and (**B**) 2004. The net embodied emissions are indicated by dots.

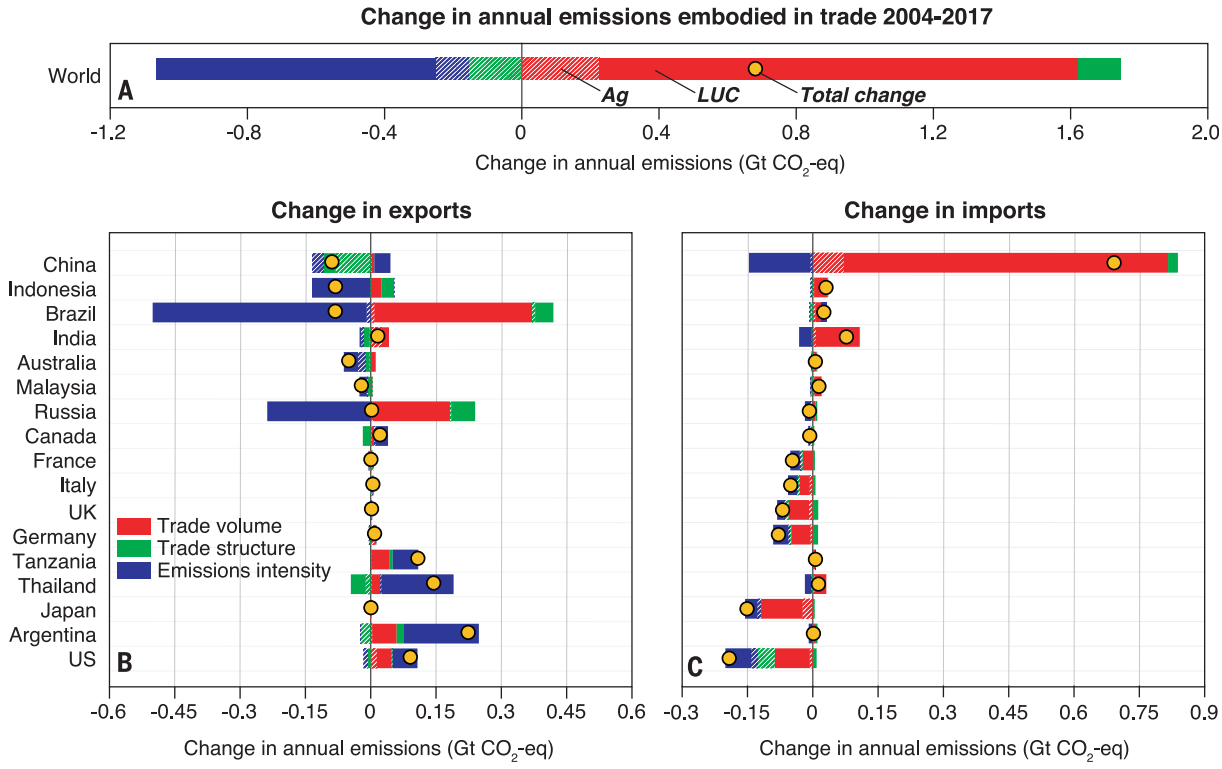


Fig. 6. Drivers of changes in land-use emissions embodied in trade. (**A** to **C**) Contributions of different factors to changes in annual land-use emissions embodied in global trade (A), and exports (B) and imports (C) of major traders from 2004 to 2017. The contributions of each factor (i.e., trade volume, trade structure, and emissions intensity; materials and methods) are shown by bars and total changes over 2004 to 2017 are shown by yellow circles. Total land-use emissions are separated into land-use change emissions (LUC; solid bars) and agricultural emissions (Ag; hatched bars).

"indirect land use change" emissions related to agricultural and forest policies in the US and Europe, particularly related to biofuels and bioenergy (e.g., Europe's decision to phase palm oil out of biofuel markets) (*9*, *43*–*45*). We find that three-quarters of land-use emissions embodied in trade are related to land-use change, which is known to also have considerable ecological impacts (*1*, *2*). If it were possible to produce the same crops by sustainable intensification (*46*–*48*), land-use emissions embodied in trade would be drastically reduced. Despite high LUC emissions, more than a quarter of agricultural land in Brazil and Argentina supports exports, suggesting that targeted zero deforestation agreements among international commodities traders could yield substantial environmental benefits. Policies similar to the Amazon Soy Moratorium (*34*) might be bolstered and expanded to include more regions and commodities such as soy from Brazil's Cerrado [which has been proposed (*35*, *49*)], palm oil from Indonesia and cocoa from Equatoria Africa. Well-crafted policies could help ensure that any additional costs associated with avoiding such land-use change would at least initially be borne by (typically more affluent) importing regions. However, proposals for Europe and the US to adopt border carbon fee adjustments to prevent carbon leakage related to these regions' climate mitigation efforts have rarely extended to land-use emissions (*50*), and care would be needed to ensure that such policies were not implemented in a regressive manner.

Our results, which identify the regions, products, and trade relationships contributing most to the transfer of land-use emissions, may help target efforts to improve the sustainability of land use and agricultural production. In particular, international trade may be used to reduce the emission intensity of agricultural products in the future, on the basis of comparative environmental advantage. For example, China's imports of soybeans from the US are less emissions-intensive than that from Brazil and Argentina (fig. S7), suggesting that global emissions might be avoided if the US were able to produce and export more soy to China (*51*). Of course, it is not quite this straightforward as a result of the dynamism of global trade networks and other environmental factors such as water resources and biodiversity, as well as the influence of various and local social, political, and economic interests (*52*). Indeed, to better support local decision making in the future, datasets and analyses of food production and trade may be extended to resolve food system dynamics at subnational or finer spatial scales; for example, shifts in the use of marginal lands. By the same token, different datasets and accounting schemes can allocate land-use emissions differently among products and across time (materials and methods) (*23*). Nonetheless, our results make the case that policies incorporating environmental externalities could help broaden the notion of comparative advantage, as well as reduce incentives for importers to offshore GHG emissions and for exporters to backslide into environmentally destructive practices. Land-use emissions are expected to be a key challenge in climate mitigation efforts as global population and food demands increase (*3*, *22*). Our results demonstrate the importance of assessing land-use emissions embodied in trade to avoid policy-related leakage and reveal targeted opportunities for international cooperation to reduce emissions and land intensities of global agricultural production.

REFERENCES AND NOTES

1. J. A. Foley *et al.*, *Science* **309**, 570–574 (2005).
2. T. Newbold *et al.*, *Nature* **520**, 45–50 (2015).
3. C. Hong *et al.*, *Nature* **589**, 554–561 (2021).
4. G. K. MacDonald *et al.*, *Bioscience* **65**, 275–289 (2015).
5. H. Haberl *et al.*, *Geogr. Tidsskr.* **109**, 119–130 (2009).
6. S. J. Davis, K. Caldeira, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 5687–5692 (2010).
7. G. P. Peters, J. C. Minx, C. L. Weber, O. Edenhofer, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 8903–8908 (2011).

8. Q. Zhang *et al*., *Nature* **543**, 705–709 (2017).
9. P. Meyfroidt, T. K. Rudel, E. F. Lambin, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 20917–20922 (2010).
10. G. Q. Chen, M. Y. Han, *Land Use Policy* **49**, 118–130 (2015).
11. Y. Yu, K. S. Feng, K. Hubacek, *Glob. Environ. Change* **23**, 1178–1186 (2013).
12. C. Dalin, Y. Wada, T. Kastner, M. J. Puma, *Nature* **543**, 700–704 (2017).
13. A. K. Chapagain, A. Y. Hoekstra, H. H. G. Savenije, *Hydrol. Earth Syst. Sci.* **10**, 455–468 (2006).
14. J. Karstensen, G. P. Peters, R. M. Andrew, *Environ. Res. Lett.* **8**, 024005 (2013).
15. D. Caro, A. LoPresti, S. J. Davis, S. Bastianoni, K. Caldeira, *Environ. Res. Lett.* **9**, 114005 (2014).
16. E. Hansis, S. J. Davis, J. Pongratz, *Global Biogeochem. Cycles* **29**, 1230–1246 (2015).
17. S. J. Davis, J. A. Burney, J. Pongratz, K. Caldeira, *Carbon Manag.* **5**, 233–245 (2014).
18. IPCC Working Group III Technical Support Unit, "Climate Change 2014: Mitigation of Climate Change" (Cambridge Univ. Press, 2014).
19. S. Roe *et al*., *Nat. Clim. Chang.* **9**, 817–828 (2019).
20. K. M. Carlson *et al*., *Nat. Clim. Chang.* **7**, 63–68 (2017).
21. P. Friedlingstein *et al*., *Earth Syst. Sci. Data* **11**, 1783–1838 (2019).
22. M. A. Clark *et al*., *Science* **370**, 705–708 (2020).
23. Materials and methods are available as supplementary materials.
24. FAO, FAOStat (2019); www.fao.org/faostat/en/.
25. T. Kastner, M. Kastner, S. Nonhebel, *Ecol. Econ.* **70**, 1032–1040 (2011).
26. P. D'Odorico, J. A. Carr, F. Laio, L. Ridolfi, S. Vandoni, *Earths Futur.* **2**, 458–469 (2014).
27. J. Poore, T. Nemecek, *Science* **360**, 987–992 (2018).
28. S. Henders, U. M. Persson, T. Kastner, *Environ. Res. Lett.* **10**, 125012 (2015).
29. F. Pendrill *et al*., *Glob. Environ. Change* **56**, 1–10 (2019).
30. N. Escobar *et al*., *Glob. Environ. Change* **62**, 102067 (2020).
31. N. T. Hoang, K. Kanemoto, *Nat. Ecol. Evol.* **5**, 845–853 (2021).
32. A. Baccini *et al*., *Nat. Clim. Chang.* **2**, 182–185 (2012).
33. M. Herrero *et al*., *Nat. Clim. Chang.* **6**, 452–461 (2016).
34. R. Heilmayr, L. L. Rausch, J. Munger, H. K. Gibbs, *Nat. Food* **1**, 801–810 (2020).
35. A. C. Soterroni *et al*., Expanding the Soy Moratorium to Brazil's Cerrado. *Sci. Adv.* **5**, eaav7336 (2019).
36. J. S. H. Lee, D. A. Miteva, K. M. Carlson, R. Heilmayr, O. Saif, *Environ. Res. Lett.* **15**, 124064 (2020).
37. J. Assunção, C. Gandour, R. Rocha, *Environ. Dev. Econ.* **20**, 697–722 (2015).
38. C. H. L. SilvaJunior *et al*., *Nat. Ecol. Evol.* **5**, 144–145 (2021).
39. IPCC, "Guidelines for National Greenhouse Gas Inventories" (IPCC, 2006).
40. European Commission, Commission Implementing Regulation (EU) 2020/1208 of 7 August 2020; https://eur-lex.europa.eu/eli/reg_impl/2020/1208/oj.
41. USEPA, Greenhouse Gas Reporting Program, https://www.epa.gov/ghgreporting.
42. C. Dalin, *Proc. Natl. Acad. Sci. U.S.A.* **109**, 5989–5994 (2012).
43. J. Fargione, J. Hill, D. Tilman, S. Polasky, P. Hawthorne, *Science* **319**, 1235–1238 (2008).
44. T. Searchinger *et al*., *Science* **319**, 1238–1240 (2008).
45. E. F. Lambin, P. Meyfroidt, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 3465–3472 (2011).
46. J. A. Burney, S. J. Davis, D. B. Lobell, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 12052–12057 (2010).
47. A. Lamb *et al*., *Nat. Clim. Chang.* **6**, 488–492 (2016).
48. D. Tilman, C. Balzer, J. Hill, B. L. Befort, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 20260–20264 (2011).
49. E. K. H. J. zu Ermgassen *et al*., *Environ. Res. Lett.* **15**, 035003 (2020).
50. J. Dumortier *et al*., *Environ. Res. Lett.* **7**, 024023 (2012).
51. R. Fuchs *et al*., *Nature* **567**, 451–454 (2019).
52. J. Lin *et al*., *Nat. Commun.* **10**, 4947 (2019).

ACKNOWLEDGMENTS

The authors thank all the developers of the GTAP Data Base, E. Hansis for making contributions to the development of the BLUE model, and J.E.M.S. Nabel for performing initial BLUE simulations. **Funding:** C.H. acknowledges support from the Scientific Research Start-up Funds (QD2021030C) from Tsinghua Shenzhen International Graduate School. J.A.B., F.C.M., and S.J.D. were supported by the US National Science Foundation and US Department of Agriculture (INFEWS grant EAR 1639318), and S.J.D. also acknowledges a grant from the ClimateWorks Foundation (22-2100 / 2202-802317943). Q.Z. was supported by the National Natural Science Foundation of China (41921005). H.Z. was supported by the National Natural Science Foundation of China (71904097). Y.L. was supported by the National Natural Science Foundation of China (72125010 and 71974186). R.B.J. acknowledges support from the Gordon and Betty Moore Foundation (Grant GBMF5439, "Advancing Understanding of the Global Methane Cycle"). **Author contributions:** C.H. and S.J.D. conceived the study. C.H., H.Z., and Y.Q. performed the analyses, with support from Q.Z., J.P., K.H., and Y.L. on datasets, and from S.J.D., Q.Z., J.A.B., J.P., F.C.M., and R.B.J. on analytical approaches. C.H. and S.J.D. led the writing with input from all coauthors. All coauthors reviewed and commented on the manuscript. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Production-based estimates of land-use emissions are available at https://sustsys.ess.uci.edu/CALUE.html. Statistics on agricultural production and land area as well as the agricultural emissions are curated by the FAO and freely available at www.fao.org/faostat/en/. The GTAP Data Base is available at https://www.gtap.agecon.purdue.edu/databases/. The supplementary materials contain our results. All data central to the main claims are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abj1572
Materials and Methods
Figs. S1 to S17
Tables S1 to S3
References (*53*–*69*)
Data S1 to S5

Submitted 23 April 2021; resubmitted 26 January 2022
Accepted 31 March 2022
10.1126/science.abj1572

CATALYSIS

The state of zinc in methanol synthesis over a Zn/ZnO/Cu(211) model catalyst

Peter Amann[1]†, Bernhard Klötzer[2], David Degerman[1], Norbert Köpfle[2]‡, Thomas Götsch[3], Patrick Lömker[4,1], Christoph Rameshan[5], Kevin Ploner[2], Djuro Bikaljevic[2], Hsin-Yi Wang[1], Markus Soldemo[1]§, Mikhail Shipilin[1], Christopher M. Goodwin[1], Jörgen Gladh[1]§, Joakim Halldin Stenlid[1]¶, Mia Börner[1], Christoph Schlueter[4], Anders Nilsson[1]*

The active chemical state of zinc (Zn) in a zinc-copper (Zn-Cu) catalyst during carbon dioxide/carbon monoxide (CO_2/CO) hydrogenation has been debated to be Zn oxide (ZnO) nanoparticles, metallic Zn, or a Zn-Cu surface alloy. We used x-ray photoelectron spectroscopy at 180 to 500 millibar to probe the nature of Zn and reaction intermediates during CO_2/CO hydrogenation over Zn/ZnO/Cu(211), where the temperature is sufficiently high for the reaction to rapidly turn over, thus creating an almost adsorbate-free surface. Tuning of the grazing incidence angle makes it possible to achieve either surface or bulk sensitivity. Hydrogenation of CO_2 gives preference to ZnO in the form of clusters or nanoparticles, whereas in pure CO a surface Zn-Cu alloy becomes more prominent. The results reveal a specific role of CO in the formation of the Zn-Cu surface alloy as an active phase that facilitates efficient CO_2 methanol synthesis.

Methanol receives much interest as an industrial chemical as well as for its potential as an energy carrier (*1*). Its gravimetric energy density is comparable to that of liquid ammonia and can easily be distributed and stored with established technologies. To date, the leading industrial catalyst for the CO/CO_2 hydrogenation toward methanol consists of a mixture of Cu, ZnO, and Al_2O_3 (*2*). Methanol synthesis is carried out between 150° and 300°C at pressures of 50 to 100 bar but can also occur at substantially lower pressures of a few millibar (*3*).

Although this catalytic process is a century old, years of intense research have aimed to determine the mechanism and active sites. The catalytic nature of the Zn in the catalyst is still highly debated (*2*), and the mechanism is not yet fully experimentally verified. The addition of ZnO/Zn goes beyond the role of a mere particle dispenser and acts as a chemically active promoter of the reaction (the ZnCu synergistic effect). There are a number of hypotheses about the chemical state of Zn—more specifically, whether it is in a metallic state as a surface alloy (*4*) or a bulk alloy (*5*), and whether small ZnO islands are present on the surface (*6*, *7*) or in bulk forms (*8*, *9*). There is also a proposal that mixed phases could coexist, either in the bulk or on the surface (*10*). Additionally, Zn diffusion and Zn spillover on

[1]Department of Physics, Stockholm University, AlbaNova University Center, 10691 Stockholm, Sweden. [2]Institute of Physical Chemistry, University of Innsbruck, Innrain 52c, 6020 Innsbruck, Austria. [3]Department of Inorganic Chemistry, Fritz Haber Institute of the Max-Planck Society, Faradayweg 4-6, 14195 Berlin, Germany. [4]Photon Science, Deutsches Elektronen-Synchrotron DESY, Notkestr. 85, 22607 Hamburg, Germany. [5]Institute of Materials Chemistry, Technische Universität Wien, Getreidemarkt 9/BC/01, 1060 Vienna, Austria.
***Corresponding author. Email: andersn@fysik.su.se**
†Present address: Scienta Omicron AB, Danmarksgatan 22, 75323 Uppsala, Sweden.
‡Present address: Plansee SE, Metallwerk-Planseestraße 71, 6600 Reutte, Austria.
§Present address: PULSE Institute, SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA.
¶Present address: SUNCAT Center for Interface Science and Catalysis, SLAC National Accelerator Laboratory, 2575 Sand Hill Road, Menlo Park, CA 94025, USA.

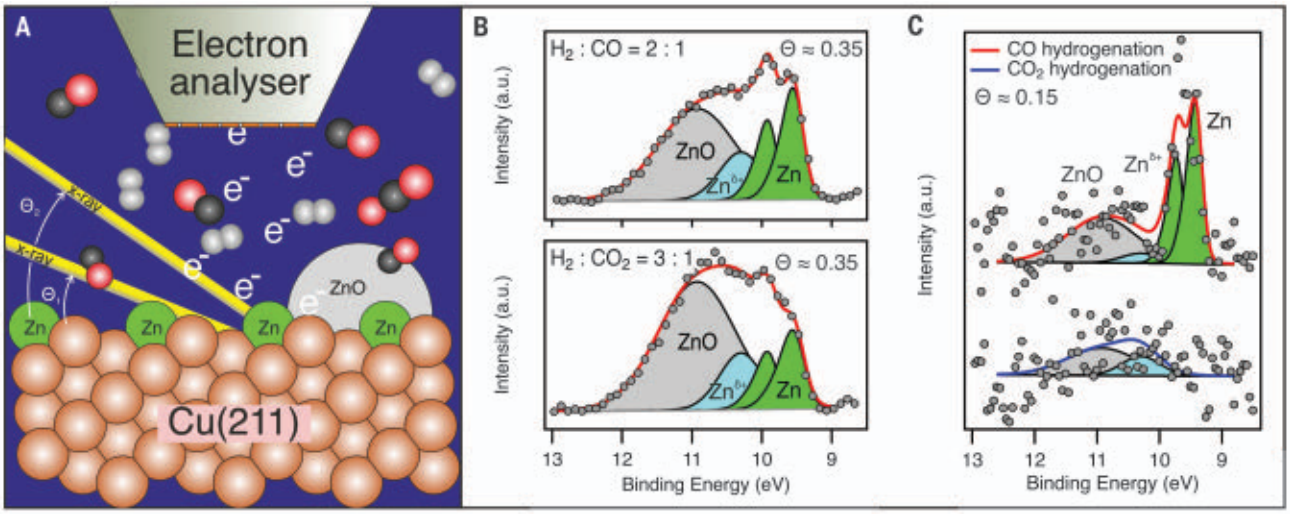


Fig. 1. Experimental setup and x-ray photoelectron spectra under different gas environments. (**A**) The surface of Zn/ZnO/Cu(211) is probed with grazing incidence x-rays at different angles (Θ_1 and Θ_2) while the surface is exposed to an elevated pressure of a gas mixture of CO, CO_2, and H_2. The surface was heated from the backside to achieve reaction conditions. (**B**) Experimental x-ray photoelectron spectra of the Zn 3d region obtained at 180 mbar, 140°C, and stoichiometric gas composition, using 4750 eV of photon energy. The experiment was conducted at ~35% surface Zn. The relative amount of ZnO to metallic Zn shows a strong dependency on gas composition. a.u., arbitrary units. (**C**) Experimental x-ray photoelectron spectra of the Zn 3d region obtained at 500 mbar, 230°C, and H_2:CO = 2.6:1 and H_2:CO_2 = 2.6:1, using 4600 eV of photon energy at a Zn coverage of ~15%. The spectra contain more noise because of the elevated pressure but exhibit more-pronounced differences between the two gas compositions.

the Cu surface have been observed at various conditions, suggesting that the system can have a strongly dynamic character (*4*, *11*). These highly divergent hypotheses result from a complex interplay between Cu and ZnO and the relative proportions of CO_2 and CO in the reactant gas and thus have sparked intense debate (*12*, *13*).

Another mechanistic issue is whether the Zn-promoted mechanism propagates through formate and methoxy species as stable intermediates with different propensities, depending on whether the reactants are CO_2, CO, or a mixture of both (*4*, *6*, *14*, *15*). It has been determined from isotope labeling experiments that the reaction proceeds more efficiently from CO_2 than CO (*16*), but it remains unresolved how the mixture of the two gases yields the highest production rate of methanol (*17*). However, the cooperative action of ZnO_x and CO_2 has also been reported to enhance the rate of methanol synthesis from CO, even when the CO_2 fraction is as low as 3% (*18*).

Experimental observation of the response from the potential active sites and surface intermediates in situ could test the various hypotheses about the active catalytic site, but such studies would need to be conducted near the operating conditions at which the reaction turns over (*19*). As the catalyst is exposed to elevated temperatures and pressures with a variable reaction mixture, it may undergo critical changes that take place only in the uppermost atomic layers. In many cases, only a few active centers at step edges are responsible for reaction propagation (*20*). Active centers and reaction intermediates often amount to minority species in a dominating macroscopic bulk system of the catalyst material together with a large macroscopic gas volume. To meet this challenge, recent studies have conducted in situ x-ray absorption spectroscopy of the Cu-Zn system at high-pressure conditions (*8*–*10*), but only with the use of bulk-sensitive detection schemes.

X-ray photoelectron spectroscopy (XPS) can be used to investigate the chemical nature of catalytic surfaces and adsorbates through core-level shifts. In particular, all aspects of the catalyst system in terms of the metallic/oxide surface, adsorbates, and gas phase can be probed under identical conditions (*21*), and well-defined peaks are often observed. The high inelastic scattering cross section of photoelectrons in the gas phase makes vacuum conditions necessary. The approach for catalysis studies using XPS has either been postreaction analysis after the reaction at high pressure or under near–ambient-pressure conditions with a differential pumping arrangement (*22*).

Near–ambient-pressure conditions in the 1-mbar regime are still limited in terms of high-pressure capability when reactions proceed at high rates (*23*). For the Cu-Zn system, the pressure has been restricted to between 0.05 and 1 mbar (*7*, *10*, *24*). Nakamura *et al.* have pointed out that postreaction studies can potentially result in misleading conclusions about the active state of Zn in methanol synthesis (*12*), in which reaction intermediates [e.g., formate (HCOO)] may decompose and oxidize the surface when the system is evacuated and the temperature is reduced. Kuld *et al.* (*3*) and Behrens *et al.* (*4*) showed that the Zn coverage dynamically responds to the chemical potential of the surrounding gas, which, for postreaction experiments, will be changed upon evacuation.

In this work, we demonstrate that a specially engineered ambient-pressure XPS experiment, based on a design with local high pressure and extreme grazing incidence of incoming x-rays, is capable of providing high surface sensitivity (*25*). This approach enables investigation of the nature of Zn and surface-adsorbed intermediates with a pressure of several hundred millibar at elevated temperatures, thereby shifting toward more-realistic conditions for methanol synthesis. To enable a direct comparison with theoretical calculations (*4*) that aimed to describe the industrial catalytic process, we selected an identical model system with a stepped Cu(211) single crystal promoted by Zn. This system exhibits superior turnover frequency relative to other more-compact surfaces such as Cu(111) and Cu(100) (*4*, *6*).

Furthermore, the high concentration of steps at the surface acts to simulate defects, as motivated by the study of Behrens *et al.* (*4*). From Zn 3d core-level shifts, we found that the nature of Zn depended on the reaction gas mixture. Incidence angle–dependent spectra showed that, in CO_2-rich conditions, ZnO was favored to exist as bulk-like particles, whereas

CO tended to generate metallic Zn that was alloyed with the Cu near the surface. From the C 1s spectra, we concluded that for CO_2 reduction there was a codominance of formate and methoxy as long-lived intermediates, whereas in CO the methoxy species dominated. The coverage of the intermediates diminished at higher temperature, indicating that the reaction was turning over. The aforementioned results support a model in which higher methanol production yield with a mixture of CO_2 and CO (*3*) is a consequence of the strong reducing ability of CO, which generates a surface that has a high density of alloyed Cu-Zn sites (*3*) and is particularly active for CO_2 reduction to methanol (*2*).

Outline of the experiment

The principle of the experiment is illustrated in Fig. 1A (*26*). The Zn/ZnO/Cu(211) surface was prepared by evaporating metallic Zn and subsequent thermal annealing, generating around 15 or 35% of surface Zn inside the vacuum chamber without exposure to air [higher annealing temperature or longer annealing time led to less Zn surface content (*26*)]. The sample surface was then placed parallel to and within 30 μm of the electron spectrometer entrance. Precleaned and premixed gases of CO, CO_2, and H_2 were directed onto the sample surface, which created a localized volume of elevated pressure that acted as a small virtual cell with rapid gas flow. Using a well-focused and low-divergence x-ray beam from beamline P22 at the Petra III synchrotron radiation facility, we probed the surface under grazing incidence conditions with a precision of ±2 μrad (*25*). Given the specific reflectivity of the Cu single crystal, the choice of incidence angles below and above the critical angle of reflection allowed us to obtain depth information (*27*). To account for the variation in the pressure and differences in photoelectron absorption in the gas phase, all spectra were normalized to the simultaneously recorded Cu $2p_{3/2}$ core-level spectra. The measurements were conducted below a photon flux threshold so that beam-induced changes could be eliminated (*26*).

Owing to its high photoionization cross section, the stronger bound 2p core level is often used to determine the oxidation state of 3d metals. Previous studies focusing on the Zn $2p_{3/2}$ core level showed a shift from 1021.2 to 1021.8 eV upon oxidation of metallic Zn to ZnO (*6*), but the inherent lifetime broadening of the Zn $2p_{3/2}$ core level renders peak deconvolution highly ambiguous. However, the valence-like Zn 3d core level has less lifetime broadening and allows deconvolution into metallic and oxidized states (*28*, *29*), but with lower photoionization cross section and less signal intensity.

The importance of in situ measurements in comparison to postreaction analysis in vacuum has been demonstrated for CO_2-rich conditions (*26*). The spectra in fig. S4 show distinctly different peak positions, with a pronounced shift from oxidized Zn to more-metallic Zn under vacuum conditions (*7*). These findings emphasize the importance of in situ investigations, and we have conducted all of our experiments at pressures between 180 and 500 mbar.

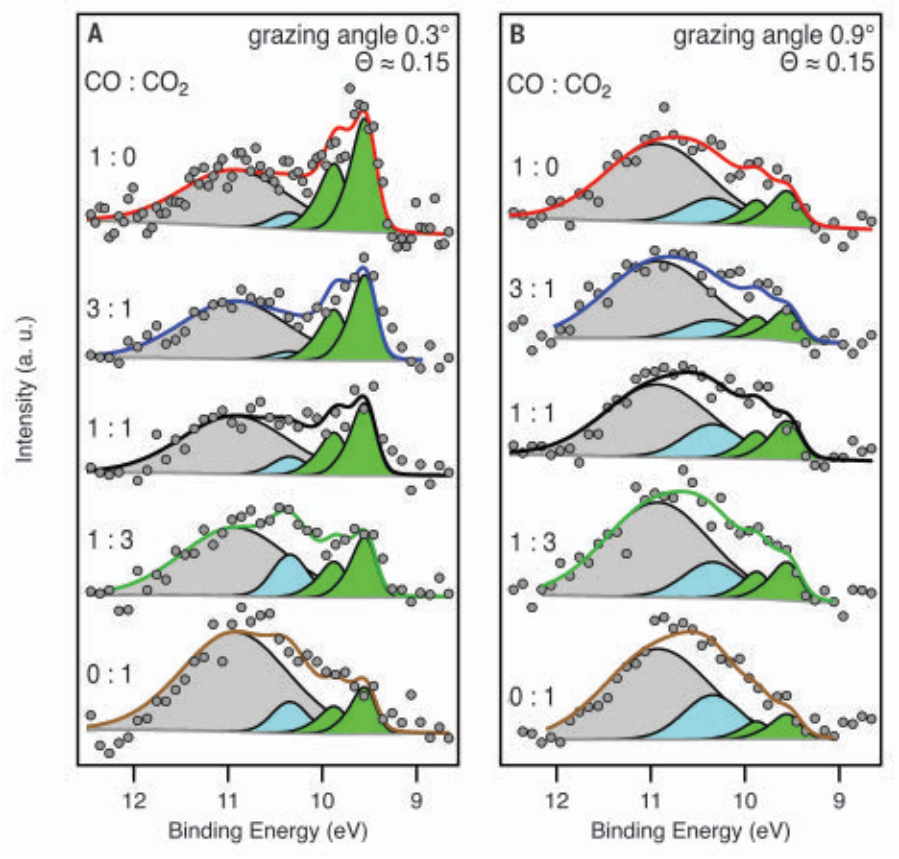


Fig. 2. Zn 3d spectral region in different gas compositions and probing depths. Spectra were collected at 180°C and ~280 mbar on the system Zn/ZnO/Cu(211) at a varying gas mixture of H_2:(CO + CO_2) = 2.6:1, with CO:CO_2 ratios indicated for ~15% surface Zn. Spectral raw data (circles) are deconvoluted by a four-peak structure. The green components at 9.53 and 9.85 eV correspond to the spin-orbit split components of metallic Zn 3d, the structure at 10.93 eV (gray) corresponds to ZnO, and the structure at 10.3 eV (light blue) corresponds to Zn in a (δ+) oxidation state. (**A**) Highly surface-sensitive experiments at a grazing incidence angle of 0.3°. A pronounced change is observed as the Zn:ZnO ratio varies. (**B**) Same experiment as in (A) but with a grazing incidence angle of 0.9°, to achieve greater bulk sensitivity. Data were accumulated consecutively, from CO:CO_2 = 1:0 to CO:CO_2 = 0:1. For each fixed-gas mixture, the angles of 0.3° and 0.9° were investigated.

Metallic Zn or ZnO depending on gas composition

In Fig. 1B we show Zn 3d spectra measured under reaction conditions of 180 mbar, 140°C, and 2:1 H_2:CO and 3:1 H_2:CO_2 gas fractions. The spectra were consistently characterized by four peaks at 10.9, 10.3, 9.8, and 9.5 eV throughout the accumulated dataset. As detailed below, the low–binding energy peaks at 9.5 and 9.8 eV (green) corresponded to the spin-orbit split states of metallic Zn $3d_{3/2}$ and Zn $3d_{5/2}$. We associated the broad structure at 10.9 eV (gray) with Zn in the +2 oxidation state, related to bulk-like ZnO. We assigned the component at ~10.3 eV (light blue) to a Zn-(δ+) oxidation state species related to Zn interacting with formate and methoxy (*4*, *30*, *31*) as well as some ZnOH. Thus, we identified the preferential states of Zn directly from the spectra under gas phase composition-dependent reaction conditions.

Under CO-rich conditions (H_2:CO = 2:1), the metallic Zn signal became more dominant (Zn:ZnO = 0.6), whereas under CO_2-rich conditions (H_2:CO_2 = 3:1), ZnO was strongly enhanced (Zn:ZnO = 0.34). When the pressure was increased to more-extreme conditions at 500 mbar and 230°C, the spectra became noisier, owing to increased gas-phase electron scattering (Fig. 1C), indicating that in pure CO metallic Zn was present almost exclusively, and in CO_2 only ZnO was observed. This observation directly indicated that CO promoted a more-reduced state enriched in metallic sites, whereas CO_2 drove the surface

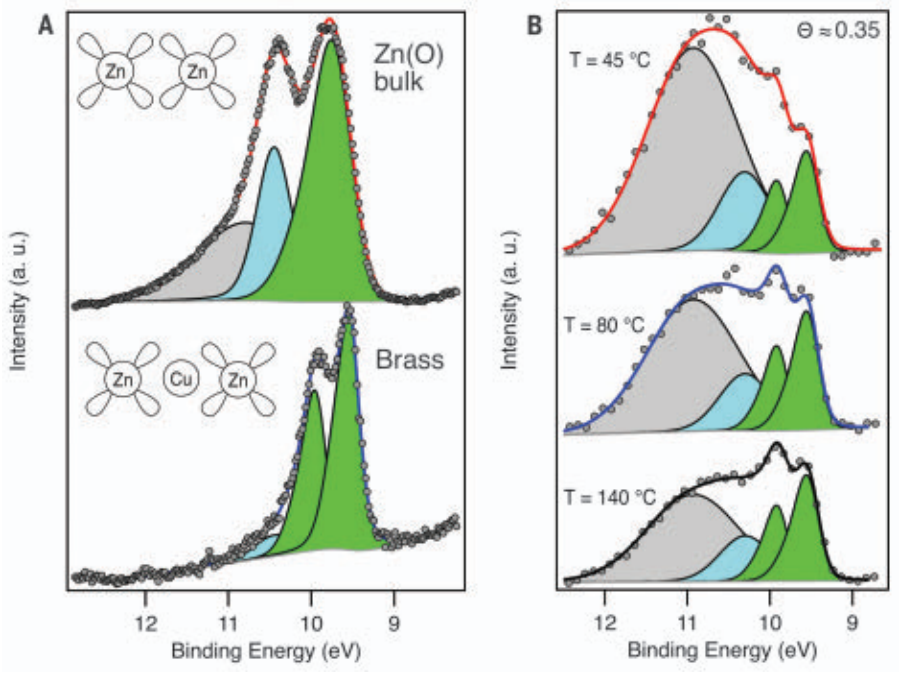


Fig. 3. Spin-orbit splitting in native polycrystalline ZnO_x and brass sample and temperature-dependent behavior of Zn on active Zn/ZnO/Cu(211). (**A**) Comparison of Zn 3d spectra measured in the hard x-ray photoelectron spectroscopy chamber at beamline P22, using a incidence angle of 5° to provide bulk-sensitive conditions. The spectra were recorded in vacuum conditions, with a photon energy of 4600 eV and at room temperature. Depending on the orbital overlap of neighboring Zn atoms, the $3d_{5/2}$-$3d_{3/2}$ spin-orbit splitting becomes apparent. (**B**) Temperature-dependent investigation of Zn/ZnO/Cu(211) with ~35% surface Zn under a grazing incidence angle of 0.6° and a stoichiometric gas ratio of H_2:CO = 2:1 at 180 mbar using the POLARIS instrument. Spectra accumulation took place in a consecutive manner from low to high temperature. Peak deconvolution and color coding are as in Fig. 2. The photon energy in this experiment was 4750 eV.

toward the fully oxidized state of Zn. We did not observe any evidence for the oxidation of Cu in either of these measurements (*26*).

The relative amount of Zn in at least two different redox states changes continuously in gas mixtures ranging from pure CO + H_2 mixture to pure CO_2 + H_2 mixture at 180°C and 280 mbar (Fig. 2). To distinguish the surface effects from bulk effects, we conducted separate measurements at two different incidence angles. At a grazing incidence angle of 0.3°, the effective probing depth was ~14 Å (*26*), which corresponded to between seven and eight layers, in accordance with the definition of interlayer distance from Gajdoš *et al.* (*32*). For 0.9° the probing depth was much greater (~53 Å) and corresponded to ~30 Cu layers. The response in peak intensity for the various gas mixtures (Fig. 2, A and B) resembles the situation discussed in connection with Fig. 1B: The peak structure under CO-rich conditions is dominated by the metallic Zn component, whereas under CO_2-rich conditions the ZnO becomes dominant—but in all cases, we find a mixed state between metallic Zn and bulk-like ZnO_x. Because higher pressure induced more electron scattering, the spectra had a lower signal-to-noise level, the spin-orbit split peaks of the Zn 3d core level became less obvious, and the intensity of the fitted light blue Zn-(δ+) component became uncertain. The situation is distinctly different at a grazing incidence angle of 0.9° (Fig. 2B). The change in envelope peak structure was barely visible as the gas composition changed, and the relative signal from metallic Zn was weaker. Because we observed a higher metallic Zn signal in comparison to ZnO_x at more grazing incidence, we concluded that the reduction of ZnO to metallic Zn occurred in close proximity to the surface or directly at the surface, whereas the main fraction of ZnO_x should represent larger assemblies containing substantial bulk ZnO contributions. The Zn-ZnO_x reduction and oxidation were reproducibly independent of the direction upon switching between CO- and CO_2-rich conditions (fig. S5) (*26*).

Distinguishing Zn-Zn and Cu-Zn alloy interactions

We used the valence character of the Zn 3d electronic state to address the nature of Zn, distinguishing ZnO_x, metallic Zn islands on the surface, or a surface Cu-Zn alloy. Because the bandwidth of Zn 3d is determined by the overlapping atomic wave functions of neighboring atoms, the 3d bandwidth became larger with increased interaction between neighboring Zn atoms. By contrast, if the Zn atoms alloyed into Cu, with d states closer to the Fermi level, the Zn 3d would become more atomic in nature. If the 3d bandwidth was larger than the $3d_{5/2}$-$3d_{3/2}$ spin-orbit splitting, only one broad 3d peak would be seen, whereas if the width was smaller than the splitting, two 3d components would be seen (*33*). The Zn 3d spectra of native polycrystalline Zn and a $Zn_{37}Cu_{63}$ brass sample measured at an incidence angle of 5° (Fig. 3A) showed ZnO and intermediate ZnO_x redox states on top of the Zn bulk metal sample, but the steeper angle allowed probing of the underlying metallic Zn at 9.76 eV. For the native polycrystalline Zn in its metallic state, we indeed saw only a broad Zn 3d metal component resembling the aforementioned large-bandwidth case, with a broad 3d peak width. However, in the brass sample, two spin-orbit components centered at 9.95 and 9.56 eV were resolved, demonstrating that the alloying with Cu could lower the 3d bandwidth of Zn. These results provided a spectral fingerprint to distinguish ZnO_x, Zn interacting with Zn, and Zn interacting with Cu in the form of an alloy.

The Zn 3d spectra in Fig. 1B showed a spin-orbit split metallic feature that was enhanced under CO-rich conditions, which was indicative of a decrease in the Zn-Zn interaction. We interpreted the lower Zn-Zn interaction as alloying with Cu, similar to the brass sample. However, we could not rule out some Zn-Zn interaction, although the extent of coordination was small relative to the Zn-Cu interaction. This alloy formation was further promoted when the temperature was raised from 45° to 140°C (Fig. 3B). With increasing temperature, we noticed a strong change in the signal intensity related to the restructuring of the surface. The ZnO, which was initially on top of the Cu surface, broke up and transformed to Zn, which further alloyed into the Cu and increased the amount of surface-alloyed Zn from 7.8% at 45°C to 9.5% at 140°C. As the total amount of Zn with respect to Cu atoms decreased with increasing temperature, it appeared that some Zn was lost into the gas phase because of the high vapor pressure of Zn metal.

By contrast, Zn atoms in the alloy were thermally more stable. Preferentially, the alloying process would start at undercoordinated Cu sites such as step edges (*4*). Metallic Zn exhibits enhanced wetting on Cu that creates an even distribution of Zn atoms on the surface to maximize Zn-Zn distance. By contrast, ZnO tends to wet less and to accumulate into ZnO clusters or nanoparticles (*34*). For CO_2 reduction on Cu(111), Kattel *et al.*, Fujitani *et al.*,

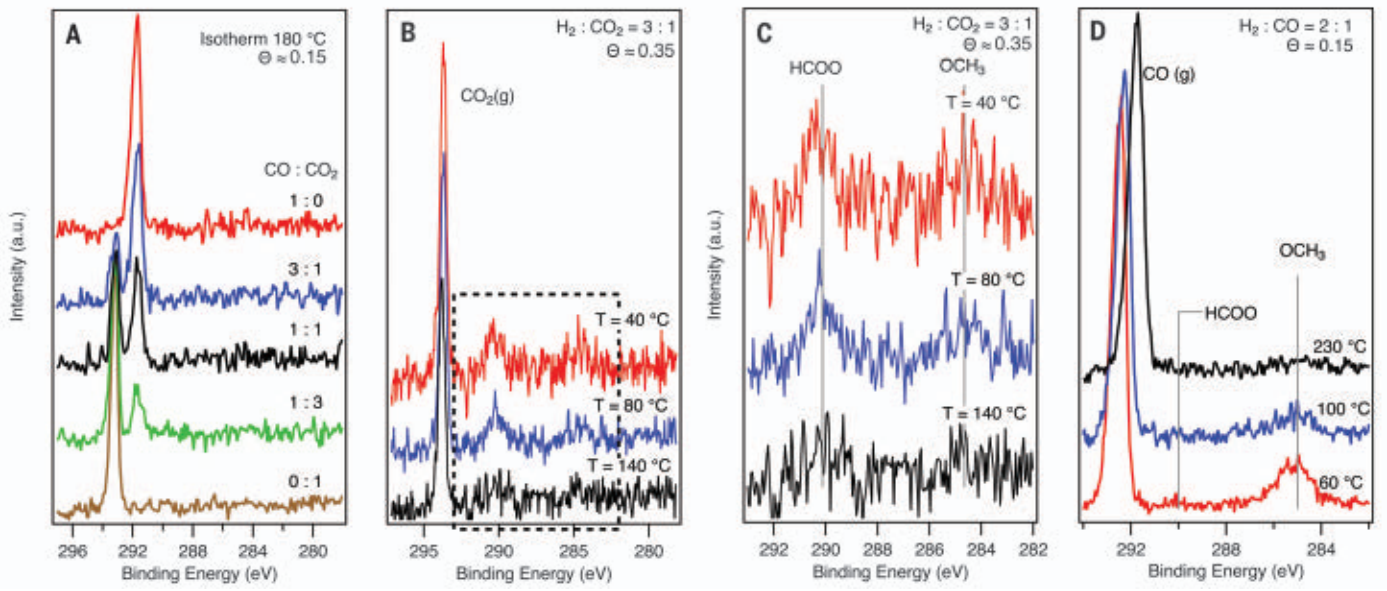


Fig. 4. Reaction intermediates probed in the C 1s region. (**A**) Surface-sensitive experiments at 180°C, ~230 mbar, and grazing incidence of 0.3°, with a varying gas mixture of H_2:(CO + CO_2) = 2.6:1. The spectra are characterized by two pronounced peaks in the region from 291 to 295 eV, corresponding to CO and CO_2 gas phase signals, and an otherwise featureless area between 278 and 291 eV. (**B**) X-ray photoelectron spectra at ~180 mbar and H_2:CO_2 = 3:1 at a grazing incidence angle of 0.6°. (**C**) Close-up view of the area surrounded by the dashed rectangle in (B), showing the presence of HCOO and OCH_3 reaction intermediates. (**D**) Spectra collected under temperature reduction in a mixture of H_2:CO = 2:1 and a pressure above ~470 mbar. For (A) and (D), data were collected with ~15% surface Zn; for (B) and (D), data were collected with ~35% surface Zn.

and Nakamura *et al.* showed that there is an optimum reactivity at a Zn coverage of ~20% of monolayer and that the chemistry can be different above or below this value (*6*, *30*, *35*). Because the reaction plotted in Fig. 3B was conducted at Zn coverage above the optimum level, some Zn may not have formed surface alloys and was readily oxidized by water vapor. However, on a stepped Cu(211) surface with more active step sites (*4*), the optimum reactivity most likely occurred at a higher Zn coverage.

Stable reaction intermediates

We next consider the reaction mechanism and the identification of stable intermediates. Under steady-state conditions, it would be possible to detect only those long-lived intermediates that can give rise to a detectable population. Density functional theory predicts deep minima in the free-energy surface, with formation of formate and methoxy in the case of CO_2 hydrogenation and formation of only methoxy in the case of CO hydrogenation (*4*, *6*, *14*). We verified that CO_2 reduction proceeds through both formate and methoxy intermediates, whereas CO reduction proceeded predominantly through methoxy. Because the O 1s region of the intermediates exhibited overlapping ZnO and ZnOH peaks, we relied on the C 1s spectral region for identification, but the corresponding O 1s spectra are also shown for completeness (fig. S14) (*26*). Many different carbon-containing molecular species potentially populate the surface, and their binding energies depend on whether the coordination is with pure Cu sites, surface-alloyed Cu-Zn sites, pristine Zn sites, ZnO_x sites, or mixed Cu-ZnO_x sites. However, we note that a given binding energy position in the spectra is consistent with the proposed intermediates.

The C 1s spectra at 180°C and 230 mbar of H_2 with CO, CO_2, or mixtures of both reactants at a 0.3° incidence angle (Fig. 4A) showed strong peaks at ~293.2 and ~291.6 eV from gas-phase CO_2 and CO, respectively. The adsorbate region between 283 and 290 eV showed no peaks, meaning that the surface coverage of any carbon-containing species was very low, consistent with a high turnover rate and a lack of long-lived intermediates. At lower temperatures (≤140°C), the turnover rate decreased and intermediates became visible. For the reaction of CO_2 with H_2, adsorbates accumulated on the surface (Fig. 4, B and C). The coverage increased with decreasing temperature, and two broad spectral structures were observed at ~290 and ~285 eV. These adsorbate peaks could be consistently assigned to formate and methoxy species (*26*). We related the C 1s formate species to some of the intensity of the Zn-(δ+) peak at 10.3 eV in Figs. 1 and 2.

For CO hydrogenation, we saw a similar trend (Fig. 4D), with a peak at ~285.2 eV that increased with decreasing temperature. Such reactivity is consistent with methoxy as the predominant intermediate (*26*). There is not a straightforward interpretation for the peaks in the carbon region, but given that multiple species can coexist at various sites, we can rule out graphitic carbon as an intermediate, as it gives rise to a much narrower peak at a lower binding energy of 285.4 eV (*36*). We stress that the interpretation of formate and methoxy on ZnO and CuZn from CO_2 and methoxy on CuZn from CO is basically consistent with the proposed mechanisms of methanol synthesis (*4*, *6*, *14*) and has been observed in low-pressure or vacuum experiments (*7*, *24*, *30*, *37*–*40*).

Inferred surface-alloyed CuZn active sites for CO_2 reduction

A schematic representation of the "working" model catalyst is illustrated in Fig. 5. In situ XPS at pressures of ≥180 mbar provided insight into the dynamic changes of the surface-near redox chemistry during reaction conditions. By focusing on the Zn 3d level, we could disentangle the effects of gas-phase composition and temperature on the surface-alloyed Zn, (partially) oxidized states of Zn, and the stability of the system Zn/ZnO/Cu(211). This complexity explained the divergent descriptions of the system that arose from the use of different reaction mixtures (*2*, *41*) and the relative lack of in situ studies. Using different incidence angles, we conclude that, in the near-surface region, Zn can convert to a surface alloyed state, whereas ZnO remains within the bulk in the form of clusters or nanoparticles on the

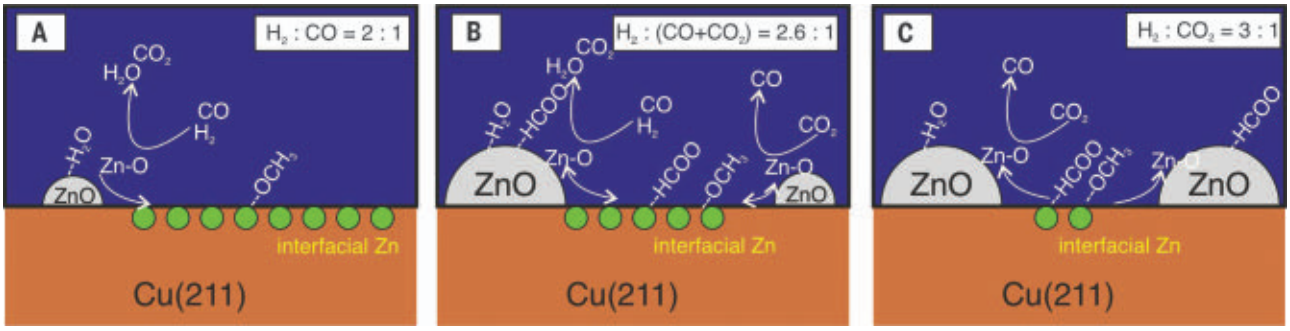


Fig. 5. Schematic illustration of the reaction mechanism. (**A** to **C**) Behavior of the Zn/ZnO/Cu(211) surface as dependent on reaction mixture and conditions, which range from CO-rich, reducing conditions in (A) to CO_2-rich, more-oxidizing conditions in (C). These illustrations show how temperature and gas mixture critically affect the ZnO ⇄ Zn equilibrium.

surface. Spectroscopically, we have demonstrated that Zn alloyed into the surface of Cu(211) to form a structure in which the Zn-Zn interaction is minimized. This observation is entirely consistent with the work of Behrens *et al.* (*4*), which suggests that Zn alloying into the Cu steps is energetically favorable.

With increasing temperature under reaction conditions, the surface gradually depleted formate and methoxy species, which is expected for a transient population of reaction intermediates. We inferred that the catalyst was very active in this state and that the observed low coverages were consistent with the findings of Kuld *et al.* (*3*). We conclude that the oxidation state of Zn was dictated by the redox chemical potential of the gas phase, rationalizing the autocatalytic behavior observed by Thrane *et al.* (*42*). Surface-alloyed Zn and ZnO_x acted as a dynamically reversible acceptor and donor, respectively, of O for the O atoms produced by the reductive activation of CO_2.

In a similar manner, Zn appeared to inhibit Cu oxidation, an effect related to the higher oxophilicity of Zn relative to Cu. Under CO-rich conditions (Fig. 5A), ZnO transformed into Zn at temperatures as low as 60°C by the formation of CO_2 and H_2O. If, alternatively, CO_2 is increasingly admixed to the reaction together with CO (Fig. 5B), it can access the active interfacial and/or metallic Zn sites and contribute to their oxidative depletion. The admixture of CO_2 to the reaction affected the local ZnO⇄Zn redox equilibrium by pushing it toward ZnO with increasing admixture of CO_2. In the absence of CO (Fig. 5C), mostly ZnO was stabilized and even enriched at 180°C. Formation of ZnO occurred at the expense of an active alloyed surface, which was effectively needed for both product-forming reactions—CO_2 reduction and direct CO hydrogenation—to proceed.

We can conclude that the most-active state is stabilized in the simultaneous presence of balanced amounts of CO, CO_2, and H_2. An enhanced CO:CO_2 ratio kept the surface more metallic, and allowed methanol formation through CO_2 to proceed more efficiently, as seen from isotope labeling experiments (*16*). The results thereby indicate that the most-active state involved an optimized surface-near abundance of redox-active surface Zn-Cu alloy sites for the reduction of CO_2, fully consistent with the proposal from Behrens *et al.* (*4*) and other studies (*3*, *11*, *43*), whereas computation studies indicate that Zn poisons CO hydrogenation (*15*).

REFERENCES AND NOTES

1. G. A. Olah, *Angew. Chem. Int. Ed.* **44**, 2636–2639 (2005).
2. J. Sehested, *J. Catal.* **371**, 368–375 (2019).
3. S. Kuld *et al.*, *Science* **352**, 969–974 (2016).
4. M. Behrens *et al.*, *Science* **336**, 893–897 (2012).
5. M. S. Spencer, *Surf. Sci.* **192**, 329–335 (1987).
6. S. Kattel, P. J. Ramírez, J. G. Chen, J. A. Rodriguez, P. Liu, *Science* **355**, 1296–1299 (2017).
7. R. M. Palomino *et al.*, *J. Phys. Chem. B* **122**, 794–800 (2018).
8. M. Zabilskiy *et al.*, *Nat. Commun.* **11**, 2409 (2020).
9. A. Beck *et al.*, *Nat. Catal.* **4**, 488–497 (2021).
10. N. J. Divins *et al.*, *Nat. Commun.* **12**, 1435 (2021).
11. A. Le Valant *et al.*, *J. Catal.* **324**, 41–49 (2015).
12. J. Nakamura *et al.*, *Science* **357**, eaan8074 (2017).
13. S. Kattel, P. J. Ramírez, J. G. Chen, J. A. Rodriguez, P. Liu, *Science* **357**, eaan8210 (2017).
14. L. C. Grabow, M. Mavrikakis, *ACS Catal.* **1**, 365–384 (2011).
15. F. Studt *et al.*, *ChemCatChem* **7**, 1105–1111 (2015).
16. G. C. Chinchen, P. J. Denny, J. R. Jennings, M. S. Spencer, K. C. Waugh, *Appl. Catal.* **36**, 1–65 (1988).
17. G. Ertl, H. Knözinger, J. Weitkamp, *Handbook of Heterogeneous Catalysis*, vols. 1 to 5 (Wiley, 2008).
18. R. Dalebout, N. L. Visser, C. E. L. Pompe, K. P. de Jong, P. E. de Jongh, *J. Catal.* **392**, 150–158 (2020).
19. A. Urakawa, *Nat. Catal.* **4**, 447–448 (2021).
20. I. Chorkendorff, J. W. Niemantsverdriet, *Concepts of Modern Catalysis and Kinetics* (Wiley, 2003).
21. M. Salmeron, R. Schlögl, *Surf. Sci. Rep.* **63**, 169–199 (2008).
22. D. E. Starr, Z. Liu, M. Hävecker, A. Knop-Gericke, H. Bluhm, *Chem. Soc. Rev.* **42**, 5833–5857 (2013).
23. J. Schnadt, J. Knudsen, N. Johansson, *J. Phys. Condens. Matter* **32**, 413003 (2020).
24. T. Koitaya *et al.*, *ACS Catal.* **9**, 4539–4550 (2019).
25. P. Amann *et al.*, *Rev. Sci. Instrum.* **90**, 103102 (2019).
26. Details of the materials and methods are available as supplementary materials.
27. L. G. Parratt, *Phys. Rev.* **95**, 359–369 (1954).
28. T. L. Barr, M. Yin, S. Varma, *J. Vac. Sci. Technol. A* **10**, 2383–2390 (1992).
29. K. Ozawa *et al.*, *Surf. Sci.* **623**, 1–5 (2014).
30. T. Fujitani, I. Nakamura, T. Uchijima, J. Nakamura, *Surf. Sci.* **383**, 285–298 (1997).
31. Y. Shiozawa, T. Koitaya, K. Mukai, S. Yoshimoto, J. Yoshinobu, *J. Chem. Phys.* **152**, 044703 (2020).
32. M. Gajdoš, A. Eichler, J. Hafner, G. Meyer, K. H. Rieder, *Phys. Rev. B* **71**, 035402 (2005).
33. M. T. Greiner *et al.*, *Nat. Chem.* **10**, 1008–1015 (2018).
34. D. Bikaljevic *et al.*, *Appl. Catal. A Gen.* **572**, 151–157 (2019).
35. J. Nakamura, Y. Choi, T. Fujitani, *Top. Catal.* **22**, 277–285 (2003).
36. A. Nikitin, L. Å. Näslund, Z. Zhang, A. Nilsson, *Surf. Sci.* **602**, 2575–2580 (2008).
37. F. Raimondi, K. Geissler, J. Wambach, A. Wokaun, *Appl. Surf. Sci.* **189**, 59–71 (2002).
38. X. Deng *et al.*, *Langmuir* **24**, 9474–9478 (2008).
39. A. V. Tarasov, F. Seitz, R. Schlögl, E. Frei, *ACS Catal.* **9**, 5537–5544 (2019).
40. I. Orozco *et al.*, *J. Phys. Chem. C* **125**, 6673–6683 (2021).
41. M. Bowker, *ChemCatChem* **11**, 4238–4246 (2019).
42. J. Thrane *et al.*, *Angew. Chem. Int. Ed.* **59**, 18189–18193 (2020).
43. D. Laudenschleger, H. Ruland, M. Muhler, *Nat. Commun.* **11**, 3898 (2020).

ACKNOWLEDGMENTS

We thank A. Gloskovskii for experimental support at the beamline and G. L. Silva Rodrigues and L. G. M. Pettersson for helpful discussions. **Funding:** This work was supported by the Swedish Research Council under grant no. 2013-8823, the Knut & Alice Wallenberg (KAW) foundation under grant no. 2016.0042, the Global Climate and Energy Project (GCEP) at Stanford University, and the Swedish Foundation for strategic research (Stiftelsen för Strategisk Forskning, SSF) under proj. nr. ITM 17-0034. We acknowledge DESY (Hamburg, Germany), a member of the Helmholtz Association HGF, for the provision of experimental facilities. Parts of this research were carried out at PETRA III using beamline P22. Beamtime was allocated for proposals I-2018014 EC and II-20190003 EC. **Author contributions:** P.A. and B.K. designed the study for beamtime I and II; P.A. and A.N. designed the study for beamtime III; P.A., B.K., and C.R. designed the experimental parameters for beamtime I, P.A. defined the parameters for beamtime II and III; P.A., B.K., C.R., N.K., and T.G. designed and implemented the evaporator; P.A., B.K., N.K., T.G., P.L., K.P., D.B., C.R., H.-Y.W., M.So., M.B., D.D., M.Sh., C.M.G., J.H.S., and J.G. conducted the experiments; P.A. analyzed the XPS data; D.D. analyzed the mass spectrometry data; P.A. created the figures; P.A., B.K., and A.N. interpreted the data; and P.A., A.N., B.K., and D.D. wrote the manuscript. **Competing interests:** The authors declare no competing interests. **Data materials availability:** All of the data necessary for evaluating the conclusions of the study are included in the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abj7747
Materials and Methods
Supplementary Text
Figs. S1 to S16
Table S1
References (*44*–*55*)
Data S1

Submitted 1 June 2021; resubmitted 9 December 2021
Accepted 29 March 2022
10.1126/science.abj7747

PHYSIOLOGY

Z-DNA binding protein 1 promotes heatstroke-induced cell death

Fangfang Yuan[1,2], Jizhen Cai[1,2], Jianfeng Wu[3,4], Yiting Tang[5], Kai Zhao[1,2], Fang Liang[1,2], Fanglin Li[1,2], Xinyu Yang[6], Zhihui He[1,2], Timothy R. Billiar[7], Haichao Wang[8], Lei Su[9], Ben Lu[1,2]*

Heatstroke is a heat stress–induced, life-threatening condition associated with circulatory failure and multiple organ dysfunctions. If global warming continues, heatstroke might become a more prominent cause of mortality worldwide, but its pathogenic mechanism is not well understood. We found that Z-DNA binding protein 1 (ZBP1), a Z–nucleic acid receptor, mediated heatstroke by triggering receptor-interacting protein kinase 3 (RIPK3)–dependent cell death. Heat stress increased the expression of ZBP1 through heat shock transcription factor 1 (HSF1) and activated ZBP1 through a mechanism independent of the nucleic acid sensing action. Deletion of ZBP1, RIPK3, or both mixed lineage kinase domain-like (MLKL) and caspase-8 decreased heat stress–induced circulatory failure, organ injury, and lethality. Thus, ZBP1 appears to have a second function that orchestrates host responses to heat stress.

Heatstroke is estimated to soon become a prominent cause of mortality worldwide (*1*, *2*). Clinically, heatstroke is characterized by extreme hyperthermia, systemic inflammatory responses, circulatory failure, bleeding and blood clotting disorder, and multiple organ dysfunctions (*1*–*3*) that result from a complex interplay between heat-related cytotoxicity, inflammation, and disseminated intravascular coagulation (DIC) (*1*, *3*). The molecular mechanisms that underlie the pathogenesis of this acute critical illness still remain largely unknown (*1*).

In *Caenorhabditis elegans*, heat stress causes pervasive necrotic cell death through calreticulin and the protease calpain—the depletion of these proteins limits heat stress–induced lethality of these nematode worms (*4*). In vertebrates, programmed necrosis is mediated by mixed lineage kinase domain-like (MLKL) or gasdermin family proteins (e.g., gasdermin D) that form pores on cytoplasmic or intracellular membranes upon activation (*5*–*9*). MLKL is activated by receptor-interacting protein kinase 3 (RIPK3)–dependent phosphorylation and executes a type of programmed necrosis, termed necroptosis (*6*, *7*). Activation of gasdermin D (GSDMD) triggers another type of programmed necrosis, termed pyroptosis (*8*, *9*). Excessive activation of GSDMD causes DIC and multiple organ dysfunctions in sepsis (*10*, *11*), an infection-induced critical illness that clinically resembles heatstroke (*1*, *10*, *11*). We investigated the possible role of programmed cell death in the pathogenesis of heatstroke and found that Z-DNA binding protein 1 (ZBP1), a Z–nucleic acid sensor (*12*–*14*), mediates the pathologic features of heatstroke by triggering RIPK3-induced activation of MLKL-dependent necroptosis and, to a lesser extent, caspase-8 (casp8)–dependent cell death.

[1]Department of Critical Care Medicine and Hematology, The 3rd Xiangya Hospital, Central South University, Changsha 410000, P.R. China. [2]Key Laboratory of Sepsis Translational Medicine of Hunan, Central South University, Changsha 410000, P.R. China. [3]Laboratory Animal Research Center, Xiamen University, Xiamen 361102, P.R. China. [4]State Key Laboratory of Cellular Stress Biology, Innovation Center for Cell Signaling Network, School of Life Sciences, Xiamen University, Xiamen 361000, P.R. China. [5]Department of Physiology, School of Basic Medical Science, Central South University, Changsha 410000, P.R. China. [6]Department of Hematology, Xiangya Hospital, Central South University, Changsha 410000, P.R. China. [7]Department of Surgery, University of Pittsburgh Medical Center, Pittsburgh, PA 15213, USA. [8]The Feinstein Institute for Medical Research, Northwell Health, Manhasset, NY 11030, USA. [9]Department of Intensive Care Unit, General Hospital of Guangzhou Command, Guangzhou 510000, P.R. China.
*Corresponding author. Email: xybenlu@csu.edu.cn

RIPK3 mediates heat stress–induced cell death and features of heatstroke

To investigate the roles of programmed cell death in the pathogenesis of heatstroke, we placed *Ripk3*$^{-/-}$, *Mlkl*$^{-/-}$, and *Mlkl*$^{-/-}$*Casp8*$^{-/-}$ mice or corresponding wild-type (WT) control mice in a temperature-controlled environmental chamber conditioned at 39°C with a relative humidity of 60 ± 5% to induce heatstroke. Markers of programmed cell death were measured over time in various tissues. The heat stress increased animals' core temperature to 43°C; induced phosphorylation of RIPK3 and MLKL in the liver, lung, and intestine within 2 hours; and increased serum concentrations of RIPK3 (Fig. 1, A and B, and fig. S1A). Heat stress also induced the cleavage of pro-casp8, pro-casp3, GSDME, and GSDMD (Fig. 1A and fig. S1A). In *Ripk3*$^{-/-}$ mice, heat stress–induced activation of programmed cell death pathways was diminished (Fig. 1A and fig. S1A). Heat stress increased serum concentrations of lactate and proinflammatory cytokines, including interleukin-1 (IL-1), IL-6, and tumor necrosis factor (TNF), and caused severe electrolyte disturbances, all of which were prevented by genetic deletion of RIPK3 (Fig. 1C and fig. S1, B and C). Heat stress also induced intravascular thrombin generation, platelet aggregation, fibrin deposition, occlusion of the microcirculation, and increase in circulating DIC markers, which were all attenuated by *Ripk3* deficiency (Fig. 1, D and E, and fig. S1, D and E). DIC promotes multiple organ injury and lethality in critical illness (*3*). In line with such clinical observations, deletion of RIPK3 attenuated heat stress–induced damage in the liver, lung, and intestine; prevented renal and pulmonary dysfunction; and prevented lethality (Fig. 1, F and G, and fig. S1, F to J). These findings indicated that RIPK3 mediates the pathologic features of heatstroke.

RIPK3 mediates necroptosis through MLKL in a kinase-dependent manner and induces apoptosis and pyroptosis through casp8 in a kinase-independent manner (*15*). To study the scaffolding function of RIPK3 in heat stress, we used *Ripk3*$^{\Delta/\Delta}$ mice that express a catalytically inactive RIPK3 resulting from the deletion of four amino acids (QWDF) in the RIPK3 kinase domain (*16*). Loss of the kinase activity of RIPK3 abrogated heat stress–induced phosphorylation of MLKL but did not affect the cleavage of pro-casp8, pro-casp3, or GSDME (Fig. 1H and fig. S2A). Inactivation of the kinase domain of RIPK1 (*17*) did not affect levels of death markers after heat stress (Fig. 1H and fig. S2A). Furthermore, deletion of both casp8 and MLKL blocked the cleavage of pro-casp3 and GSMDE and decreased the cleavage of GSDMD (Fig. 1A).

Next, we determined the kinase-dependent and kinase-independent roles of RIPK3 in heat stress–associated lethality. Deletion of RIPK3 achieved almost full protection, whereas loss of the kinase activity of RIPK3 rescued 70% of mice after lethal heat stress (Fig. 1I). In line with these observations, deletion of MLKL attenuated organ injury and rescued 75% of mice, whereas deletion of both MLKL and casp8 allowed all mice to survive heat stress (Fig. 1I). Deficiency of GSDMD conferred relatively minor protection (fig. S2B). Heat stress increased the core temperature of *Ripk3*$^{-/-}$ and *Mlkl*$^{-/-}$*Casp8*$^{-/-}$ mice to 43°C as it did in WT mice (fig. S2C), excluding the possibility that the increased survival might be the result of lower body temperature. Thus, these findings support a role for RIPK3-dependent cell death in heat stress.

Heat stress triggers cell death through activating RIPK3 in vitro

We examined whether heat stress alone could activate RIPK3-dependent cell death. Exposure of cultured L929 mouse fibroblasts to 43°C for 2 hours or 42°C for 6 hours induced the phosphorylation of RIPK3 and MLKL within 2 hours and the cleavage of pro-casp8, pro-casp3, and GSDME by 6 hours after heat stress exposure (Fig. 2A). Deletion of RIPK3 using CRISPR-CAS9 blocked heat stress–induced phosphorylation of RIPK3 and MLKL and the

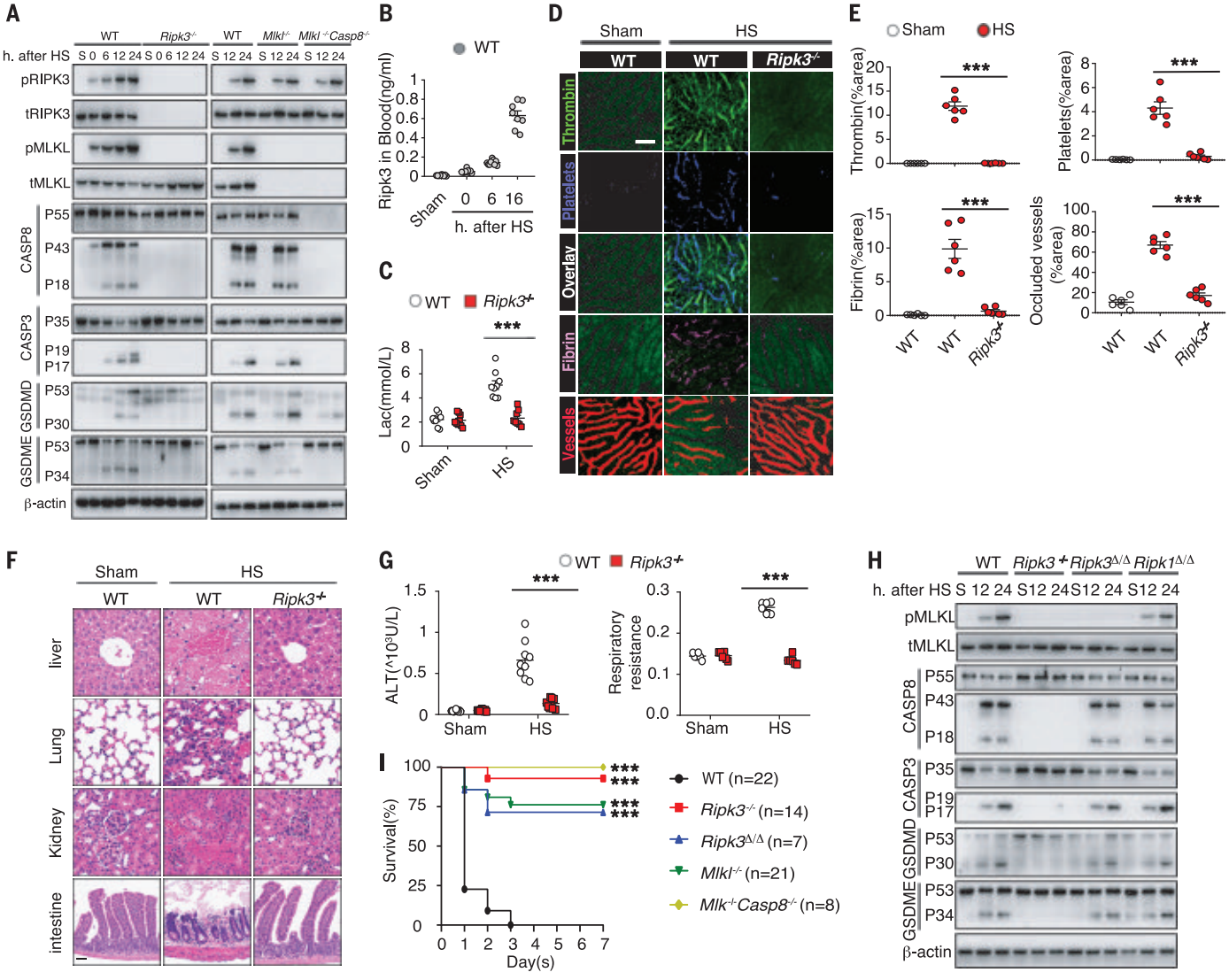
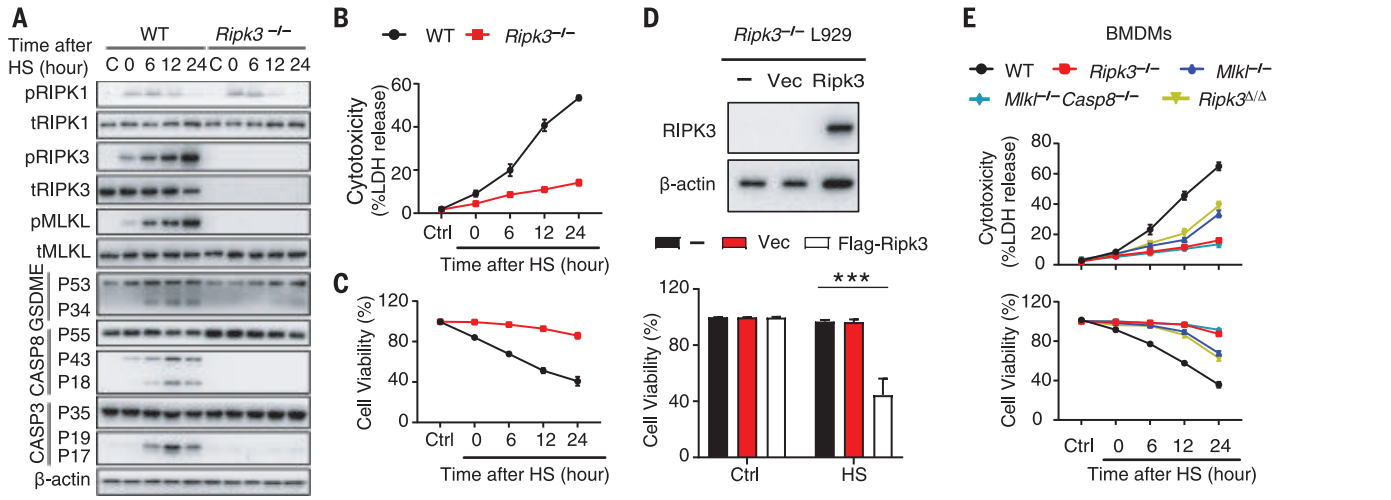


Fig. 1. RIPK3 mediates heat stress–induced cell death and features of heatstroke. (**A**) Western-blot analysis of the quantity of phosphorylated (p) RIPK3 and MLKL, total RIPK3 (tRIPK3), total MLKL (tMLKL), pro- (P55) and cleaved (P18) CASP8, pro- (P35) and cleaved (P19 and P17) CASP3, pro- (P53) and activated (P30) GSDMD, and pro- (P53) and activated (P34) GSDME in the livers of mice of indicated genotypes at the indicated time points after heat stress (HS). $n = 3$ independent biological repeats. h., hours. (**B** and **C**) Plasma concentrations of RIPK3 (B) and lactate (Lac) (C) in mice 16 hours after heat stress. (**D**) Representative images were acquired in the liver microvasculature of mice by spinning disk confocal intravital microscopy (SD-IVM) 12 hours after heat stress. Thrombin activation (green), platelet adhesion (blue), fibrin deposition (dark red), and circulating albumin (red) are shown. $n = 6$ repeats per genotype. Scale bar, 50 μm. (**E**) The fluorescence intensity was quantified by ImageJ software. (**F**) Representative images of hematoxylin and eosin (H&E) staining of different organs from WT or *Ripk3*$^{-/-}$ mice 16 hours after heat stress. Original magnification: H&E 400X. $n = 4$ repeats per genotype. (**G**) Serum alanine aminotransferase (ALT) activity and respiratory resistance were measured in WT or *Ripk3*$^{-/-}$ mice 16 hours after heat stress. (**H**) Quantity of indicated proteins in the livers of mice of indicated genotypes at 12 or 24 hours after heat stress. $n = 3$ independent biological repeats. (**I**) Survival analysis of genetic background–, age-, and sex-matched mice of indicated genotypes subjected to heat stress. Data were pooled from at least two independent experiments. Circles represent individual mice. Error bars indicate ±SEMs. ***$P < 0.001$. Statistics are by one-way analysis of variance (ANOVA) (B), two-way ANOVA test [(C), (E), and (G)], or survival curve comparison [log-rank (Mantel-Cox) test] (I).

cleavage of pro-casp8, pro-casp3, and GSDME (Fig. 2A). These events were associated with a reduction of cell death as judged by lactate dehydrogenase (LDH) release, loss of cell membrane integrity, and increased cell membrane permeability (Fig. 2, B and C, and fig. S3, A to C). Heat stress promoted the interaction between RIPK3 and MLKL in WT but not *Ripk3*-deficient L929 cells (fig. S3, D and E). Depletion of RIPK3 with short hairpin RNA (shRNA) inhibited heat stress–induced cell death (fig. S3F), whereas reconstitution of the expression of RIPK3 restored the capacity of *Ripk3*-deficient L929 cells to undergo cell death in response to heat stress (Fig. 2D). Similar observations were made using bone marrow–derived macrophages (BMDMs) and peritoneal macrophages (PMs) from WT or *Ripk3*$^{-/-}$ mice (Fig. 2, E and F, and fig. S4, A to E). Although MLKL deficiency protected cells from heat stress–induced necroptosis within 12 hours, delayed cell death occurred 24 hours after heat stress exposure in *Mlkl*$^{-/-}$ but not *Ripk3*$^{-/-}$ BMDMs (Fig. 2E). Because RIPK3 can mediate cell death through its scaffolding function and casp8 (*15*), we thought heat stress might induce RIPK3-dependent cell death through both MLKL and casp8.

To test this, *WT*, *Ripk3*$^{\Delta/\Delta}$, *Mlkl*$^{-/-}$, and *Mlkl*$^{-/-}$ *Casp8*$^{-/-}$ BMDMs were exposed to heat stress.



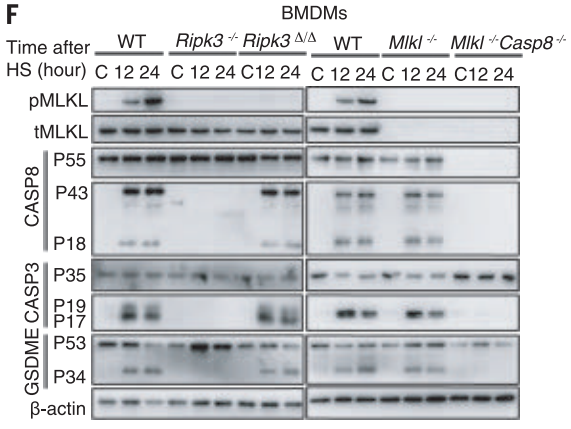
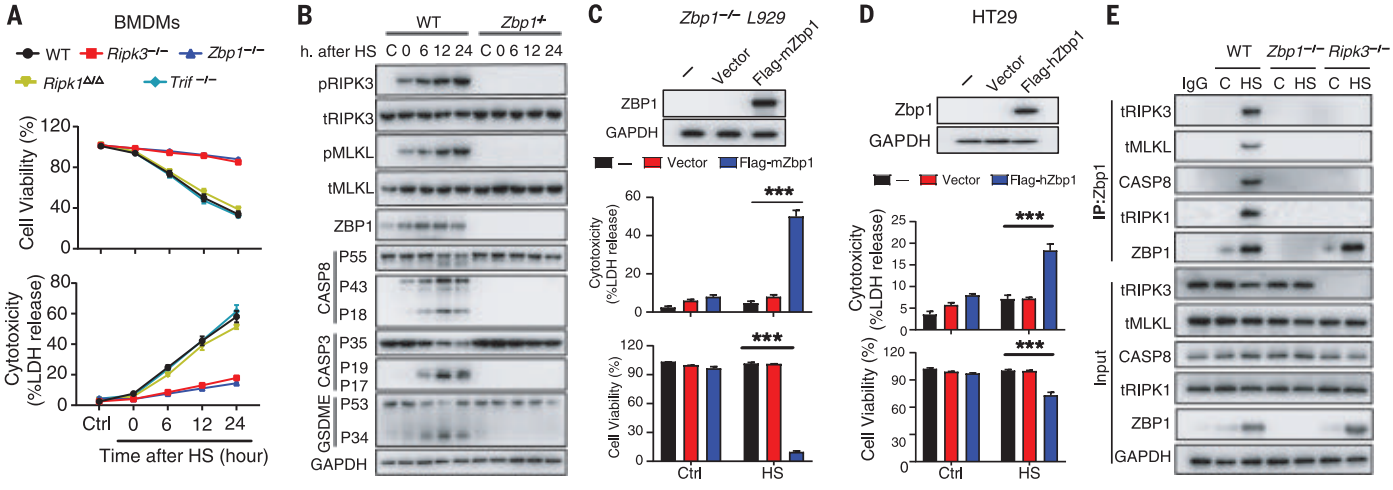


Fig. 2. Heat stress triggers RIPK3-dependent cell death. (**A**) Western-blot analysis of the quantity of indicated proteins in WT or *Ripk3*$^{-/-}$ L929 cells at the indicated time points after heat stress (43°C, 2 hours). *n* = 3 independent biological repeats. (**B** and **C**) LDH release (B) and cell viability (C) of WT or *Ripk3*$^{-/-}$ L929 cells at the indicated time points after heat stress. (**D**) *Ripk3*$^{-/-}$ L929 cells were infected with lentivirus containing blank vector (Vec) or vector encoding RIPK3. Cell viability was determined at 6 hours after heat stress. (**E**) LDH release and cell viability of BMDMs of indicated genotypes at the indicated time points after heat stress. (**F**) Quantity of indicated proteins in BMDMs of indicated genotypes after heat stress. *n* = 3 independent biological repeats. Error bars indicate ±SEMs of three independent experiments. ***$P < 0.001$. Statistics by two-way ANOVA test.



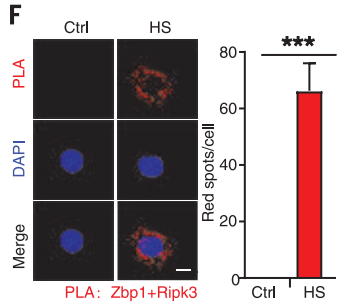


Fig. 3. Heat stress activates RIPK3 through ZBP1. (**A**) LDH release and cell viability of WT and *Zbp1*$^{-/-}$ BMDMs at the indicated time points after heat stress. (**B**) Western-blot analysis of the quantity of indicated proteins in WT and *Zbp1*$^{-/-}$ BMDMs at the indicated time points after heat stress. *n* = 3 independent biological repeats. GAPDH, glyceraldehyde-3-phosphate dehydrogenase. (**C** and **D**) *Zbp1*$^{-/-}$ L929 (mouse) (C) or HT29 cells (human) (D) were infected with lentivirus containing blank vector or vector encoding *Zbp1* for 36 hours. Cell viability and LDH release were measured at 6 hours after heat stress. (**E**) ZBP1 immunoprecipitates (IPs) in WT, *Ripk3*$^{-/-}$, or *Zbp1*$^{-/-}$ L929 cells at 6 hours after heat stress and probed for the complex. *n* = 3 independent biological repeats. IgG, immunoglobulin G. (**F**) The physical associations between ZBP1 and RIPK3 were detected as red spots using proximity ligation assay (PLA) in WT L929 cells at 6 hours after heat stress. Scale bar, 10 μm. DAPI, 4′,6-diamidino-2-phenylindole. Error bars indicate ±SEMs of three independent experiments. ***$P < 0.001$. Statistics by two-way ANOVA test [(A), (C), and (D)] or two-tailed *t* test (F).

The loss of RIPK3 kinase activity prevented the phosphorylation of MLKL but not the cleavage of casp8, casp3, and GSDME (Fig. 2F). Deletion of both MLKL and casp8 almost completely blocked the cleavage of casp3 and GSDME (Fig. 2F). Although ablation of the kinase activity of RIPK3 or deletion of MLKL prevented LDH release and cell death at 6 hours after heat stress, BMDMs from *Ripk3*$^{\Delta/\Delta}$ or *Mlkl*$^{-/-}$ mice released LDH and underwent cell death 24 hours after heat stress exposure (Fig. 2E). Delayed cell death and LDH release were prevented by deletion of RIPK3 or both MLKL and casp8 (Fig. 2E). Depletion of casp8 with shRNA or pharmacological inhibition of casp8 by Z-IEDT-FMK reduced heat stress–induced cell death in *Mlkl*-deficient L929 cells (fig. S5, A to C). Further, heat stress transiently (within 12 hours) increased the expression of cellular FLICE-like inhibitory protein (cFLIP) (fig. S5D), which inhibits casp8 activation and associated cell death (*18*). Thus, heat stress triggers cell death through RIPK3-dependent activation of MLKL and casp8.

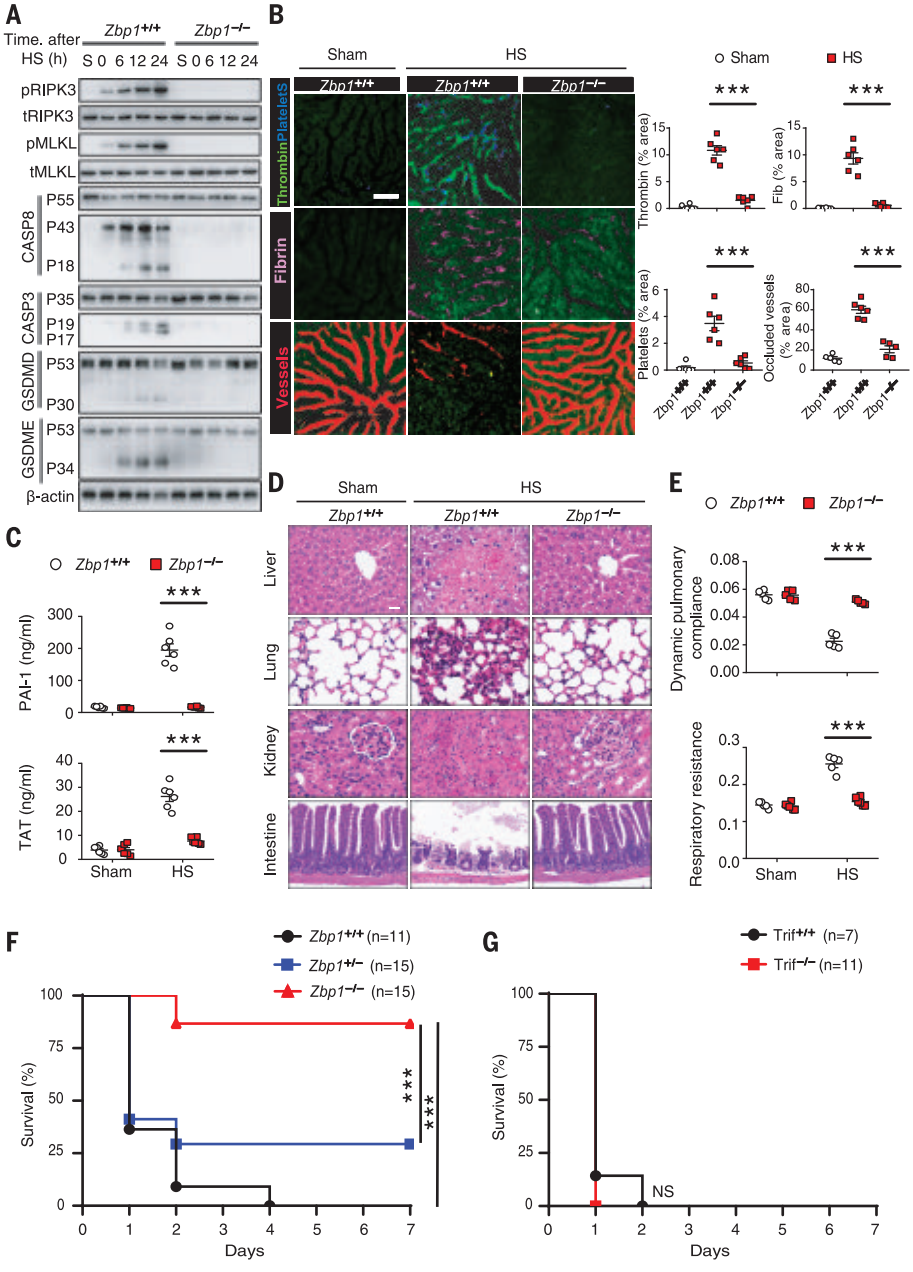


Fig. 4. ZBP1 mediates heat stress–induced cell death and features of heatstroke. (**A**) Western-blot analysis of the quantity of indicated proteins in the livers of *Zbp1*$^{+/+}$ and *Zbp1*$^{-/-}$ mice at the indicated time points after heat stress. *n* = 2 independent biological repeats. (**B**) Representative images were acquired in the liver microvasculature by SD-IVM 12 hours after heat stress. Scale bar, 50 μm. The fluorescence intensity was quantified by ImageJ software. Circles represent individual mice. Fib, fibrin. (**C**) Plasma concentrations of PAI-1 and TAT complex in *Zbp1*$^{+/+}$ and *Zbp1*$^{-/-}$ mice 16 hours after heat stress. (**D**) Representative images of H&E staining different organs from mice of indicated genotypes 16 hours after heat stress. *n* = 6 repeats per genotype. (**E**) Respiratory function in *Zbp1*$^{+/+}$ and *Zbp1*$^{-/-}$ mice at 16 hours after heat stress. (**F** and **G**) Survival analysis of mice of indicated genotypes subjected to heat stress. Data were pooled from at least two independent experiments. Circles represent individual mice. Error bars indicate ±SEMs. NS, not significant ($P \geq 0.05$); ***$P <$ 0.001. Statistics by one-way ANOVA (B), two-way ANOVA test [(C) and (E)], and survival curve comparison [log-rank (Mantel-Cox) test] [(F) and (G)].

Heat stress activates RIPK3 through ZBP1

RIPK1, Toll/interleukin-1 receptor domain–containing adapter-inducing interferon-β (TRIF), and ZBP1 are RIP homotypic interaction motif (RHIM)–containing proteins that can interact with and activate RIPK3 (*5*, *19*, *20*). We therefore investigated whether these RIPK3-interacting proteins are required for heat stress–induced cell death. Deletion of TRIF, mutation of the kinase domain of RIPK1 (*RIPK1*$^{\Delta/\Delta}$), or inhibition of RIPK1 by necrostain-1 did not affect heat stress–induced cell death in BMDMs (Fig. 3A and fig. S6, A and B). RIPK1 deficiency did not significantly inhibit heat stress–induced cell death in L929 cells (fig. S6A). By contrast, deletion of ZBP1 abrogated the phosphorylation of RIPK3 and MLKL; the cleavage of casp8, casp3, and GSDME; and cell death after heat stress (Fig. 3, A and B). These findings were confirmed in L929 cells (fig. S6, C and D). Restoration of ZBP1 expression restored the capacity of *Zbp1*-deficient cells to undergo cell death after heat stress (Fig. 3C). Human HT-29, a colon cancer cell line that express RIPK3 and RIPK1 but not ZBP1, also failed to undergo heat stress–induced cell death (Fig. 3D). Expression of exogenous human ZBP1 rendered HT-29 cells susceptible to heat stress–induced cell death (Fig. 3D). Furthermore, heat stress induced the interaction between ZBP1 and RIPK3 (Fig. 3, E and F). These data demonstrate that heat stress activates RIPK3 through ZBP1.

ZBP1 mediates the pathologic features of heatstroke

We determined whether ZBP1 activates RIPK3 and mediates heatstroke after heat stress by exposing *Zbp1*$^{-/-}$ mice and their WT littermates to heat stress. Deletion of ZBP1 blocked the heat stress–induced phosphorylation of RIPK3 and MLKL as well as the cleavage of CASP8, CASP3, GSDMD, and GSDME in the liver and intestine (Fig. 4A and fig. S7A). Loss of ZBP1 prevented heat stress–induced DIC, systemic inflammatory responses, circulatory failure, multiple organ injury, and lethality,

as similarly observed with RIPK3 deficiency (Fig. 4, B to F, and fig. S7, B to F). By contrast, deletion of TRIF or mutation of the RIPK1 kinase domain failed to protect mice against lethal heat stress (Fig. 4G and fig. S7G). Heat stress increased the core temperature of *Zbp1*$^{-/-}$ mice to 43°C as it did in WT mice (fig. S7H), excluding the possibility that the increased survival of *Zbp1*$^{-/-}$ mice might be a result of a lower body temperature. These data demonstrate that ZBP1 mediates heat stress–induced RIPK3 activation and the pathologic features of heatstroke.

Heat stress increases the expression of ZBP1 through HSF1

ZBP1 is an interferon-inducible factor (*21*), and heat stress up-regulated the expression of ZBP1 in L929 cells and mouse macrophages in a manner similar to that of type 1 interferon (Fig. 5, A to D, and fig. S8, A and B). Furthermore, heat stress stimulated ZBP1 transcription in the lung, liver, kidney, and intestine (Fig. 5E and fig. S8C). We used the bioinformatics tool Jaspar to analyze the promoter region of ZBP1 and identified a predicted binding site for heat shock transcription factor 1 (HSF1) (Fig. 5F), which is activated by heat stress (*22*). Deletion of the putative HSF1 binding site prevented the increase in transcriptional activation through the ZBP1 promoter after heat stress (Fig. 5F). Heat stress enhanced HSF1 activation and occupancy at the HSF1 binding site in the ZBP1 promoter (Fig. 5G and fig. S8, D to F). Deletion of HSF1 inhibited a heat stress–induced increase in ZBP1 expression and cell death (Fig. 5H, fig. S8G, and fig. S9A). In response to heat stress, HSF1 regulated the expression of heat shock proteins (HSPs) (fig. S9B) (*22*). HSP90 enhances TNF-induced necroptosis through the promotion of MLKL oligomerization and activation (*23*, *24*). However, depletion of HSP90 did not affect heat stress–induced phosphorylation of RIPK3 or MLKL and cell death (fig. S9, C and D). This discrepancy might be because of the different stimuli used (TNF versus heat stress). These data establish that heat stress increases the expression of ZBP1 through HSF1.

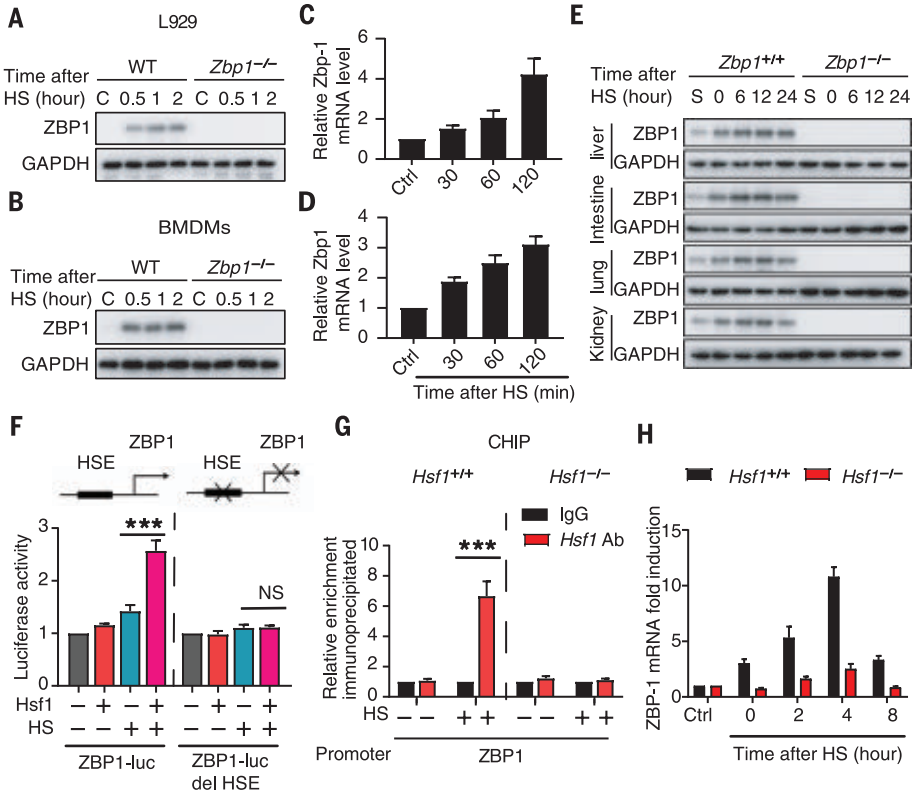


Fig. 5. Heat stress increases the expression of ZBP1 through HSF1. (**A** and **B**) Western-blot analysis of the ZBP1 expression in L929 cells (A) or BMDMs (B) subjected to heat stress for the indicated time. *n* = 3 independent biological repeats. (**C** and **D**) Quantitative real-time PCR (qRT–PCR) analysis of ZBP1 mRNA in L929 cells (C) or BMDMs (D) subjected to heat stress for the indicated time, presented relative to the quantity of GAPDH mRNA. (**E**) Western-blot analysis of ZBP1 expression in multiple organs of *Zbp1*$^{+/+}$ or *Zbp1*$^{-/-}$ mice at the indicated time points after heat stress. *n* = 3 independent biological repeats. (**F**) The transcriptional activity of WT ZBP1 promoter (ZBP1-luc) or ZBP1 promoter with a deletion of the putative heat shock element (HSE) site (ZBP1-luc del HSE) in 293T cells overexpressing HSF1 after heat stress was determined by the luciferase activity in cell lysates. (**G**) Chromatin immunoprecipitation (ChIP) assay in PMs to assess HSF1 binding at the putative HSE site in the ZBP1 promoter after heat stress. (**H**) qRT–PCR analysis of ZBP1 mRNA expression in *Hsf1*$^{+/+}$ or *Hsf1*$^{-/-}$ BMDMs at the indicated time points after heat stress. Error bars indicate ±SEMs of three independent experiments. NS ($P \geq 0.05$); ***$P < 0.001$. Statistics by two-way ANOVA test.

Z–nucleic acid sensing is dispensable for heat stress–induced ZBP1 activation

Because expression of ZBP1 by itself is insufficient to induce cell death (Fig. 3, C and D), we next investigated the mechanisms by which heat stress promotes ZBP1 activation. ZBP1 is activated by virus-derived or endogenous Z–nucleic acids during development, viral infection, and in other diseases through its Zα domain (*12*–*14*, *19*–*21*). We generated genetically modified L929 cells expressing intact ZBP1 or ZBP1 mutants that either lacked the Zα, Zα1, or Zα2 domain (ΔZα, ΔZα1, or ΔZα2) or that contained a point mutation within the Zα2 domain (Zα2 mut) that prevents Z–nucleic acid sensing (Fig. 6A). However, heat stress still triggered a ZBP1-RIPK3 interaction, RIPK3 and MLKL phosphorylation, and cell death in L929 cells expressing these ZBP1 mutants (Fig. 6, B to E). ZBP1 contains a C-terminal domain and an RHIM domain (*21*). The RHIM but not C-terminal domain was indispensable for heat stress–induced ZBP1-RIPK3 interaction, RIPK3 and MLKL phosphorylation, and cell death (Fig. 6, B to E). These findings indicate that heat stress activated ZBP1 through its RHIM domain independent of Z–nucleic acid sensing.

Heat stress promotes aggregation of ZBP1 fusion proteins

To further study how heat stress activates ZBP1, we generated HEK 293T cells transfected with plasmids that express green fluorescent protein (GFP)–tagged ZBP1. Exposure of cells to 43°C led to ZBP1-GFP aggregation into puncta within the cytosol (fig. S10A). The Zα domain was dispensable for heat stress–induced ZBP1-GFP aggregation (fig. S10A). Lack of the Zα2 domain or point mutation within the Zα2 domain did not affect ZBP1 aggregation during heat stress (fig. S10A). Immunoblots under nonreducing conditions revealed that heat stress induced the aggregation of endogenous ZBP1 (fig. S10B). Using cells expressing Flag- or Myc-tagged ZBP1 or ZBP mutants, we confirmed that exposure of cells to heat stress increased the aggregation of ZBP1 (fig. S10, C and D). The Zα domain, the C-terminal domain, and the Z–nucleic acid sensing were dispensable for ZBP1 aggregation and cell death (fig. S10, C to G).

ZBP1 contains two RHIM domains, referred to as RHIM-A and RHIM-B (*21*). Point mutation within the RHIM-A domain prevented

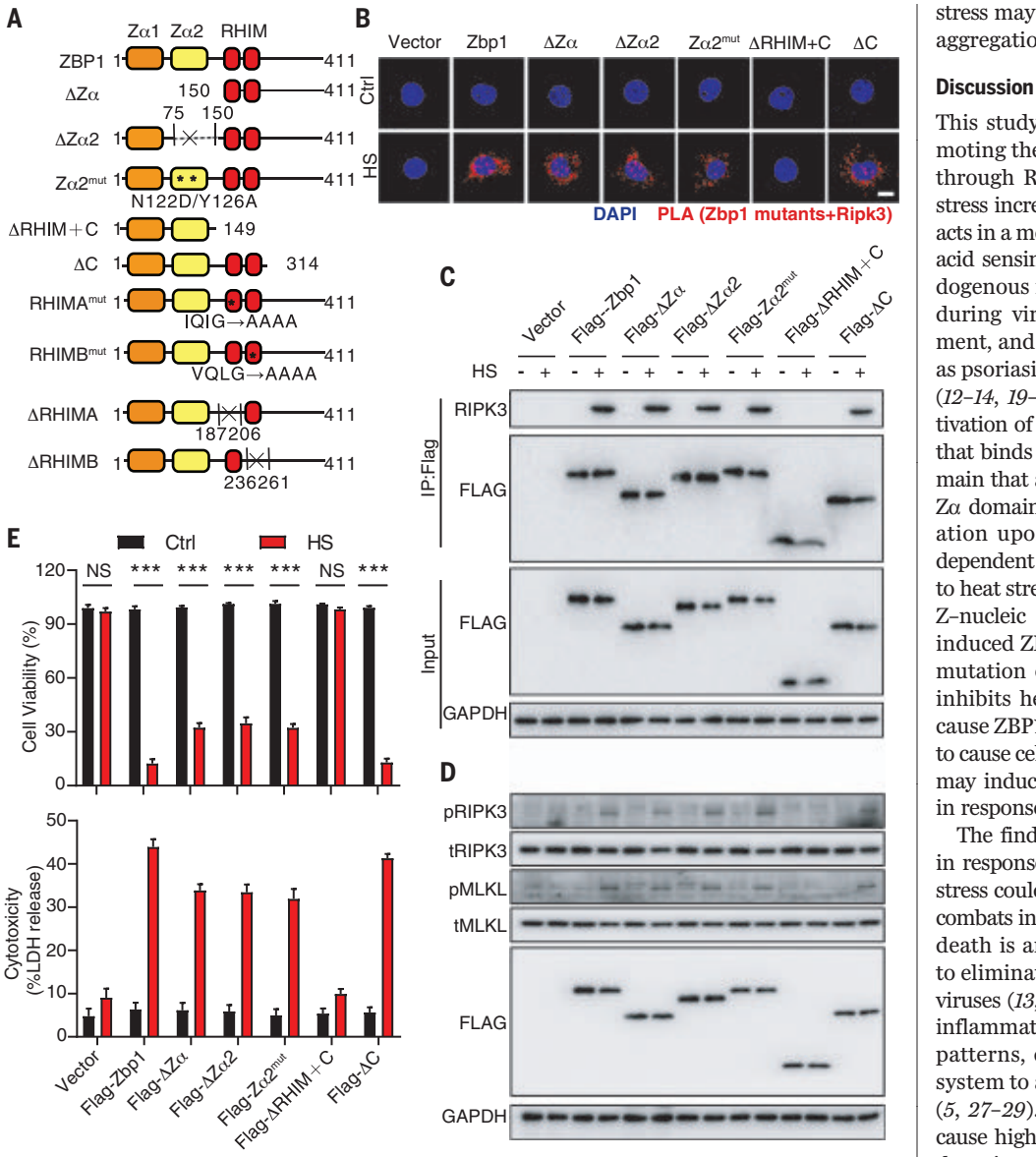


Fig. 6. Z–nucleic acid sensing is dispensable for heat stress–induced ZBP1 activation. (**A**) Schematic representation of full-length ZBP1 or indicated mutants. (B to E) *Zbp1*$^{-/-}$ L929 cells were infected with lentivirus expressing flag-tagged ZBP1 and its truncation mutants and then were subjected to heat stress or not (ctrl) at 36 hours after infection. n = 3 independent biological repeats. (**B**) The physical association between RIPK3 and ZBP1 or ZBP1 mutants was detected by PLA. (**C**) Coimmunoprecipitation assay to assess the interactions between RIPK3 and ZBP1 or its truncation mutants. Whole-cell extract (5% input) was examined in parallel for Flag-tagged ZBP1 or its mutants. (**D**) Western-blot analysis of the quantity of pRIPK3 and pMLKL at 6 hours after heat stress. (**E**) LDH release and cell viability at 6 hours after heat stress. Error bars indicate ±SEMs of three independent experiments. NS ($P \geq 0.05$); ***$P < 0.001$. Statistics by two-way ANOVA test.

heat stress–induced aggregation of ZBP1 fusion proteins (fig. S10, A to F). Reconstitution of *Zbp1*$^{-/-}$ L929 cells with WT ZBP1 and mutant ZBP1 with RHIM-B deletion or mutation, but not mutant ZBP1 with RHIM-A deletion or mutation, induced the aggregation of ZBP1 fusion proteins, the phosphorylation of MLKL, the cleavage of casp8, and cell death after heat stress (fig. S10, E to G). Aggregation of ZBP1 fusion proteins occurred as early as 30 min after heat stress, which was followed by RIPK3 recruitment, phosphorylation of MLKL, and cleavage of casp8 (fig. S10F). These events were all blocked by deletion or point mutation of the RHIM-A domain (fig. S10F). Using L929 cells expressing hormone-binding domain G521R mutant (HBD*) fused ZBP1 proteins, we observed that treatment of such cells with the HBD dimerizer 4-hydroxytamxifen (4-OHT) caused the aggregation of HBD-ZBP1 independent of heat stress (fig. S10H) and triggered cell death (fig. S10I). Thus, heat stress may promote cell death through ZBP1 aggregation.

Discussion

This study identifies a role of ZBP1 in promoting the pathologic features of heatstroke through RIPK3-dependent cell death. Heat stress increases the expression of ZBP1, which acts in a mechanism independent of Z–nucleic acid sensing. ZBP1 binds virus-derived or endogenous retrovirus-derived Z–nucleic acids during viral infection, embryonic development, and the pathogenesis of diseases such as psoriasis and inflammatory bowel disease (*12*–*14*, *19*–*21*, *25*). In these scenarios, the activation of ZBP1 requires both the Zα domain that binds Z–nucleic acids and the RHIM domain that activates RIPK3. We found that the Zα domain is dispensable for the ZBP1 activation upon heat stress. RHIM-A domain–dependent aggregation of ZBP1 may contribute to heat stress–induced cell death. Endogenous Z–nucleic acids might enhance heat stress–induced ZBP1 activation because deletion or mutation of the Zα or Zα2 domain slightly inhibits heat stress–induced cell death. Because ZBP1 expression by itself is insufficient to cause cell death, it is likely that other factors may induce ZBP1 aggregation and activation in response to heat stress.

The finding that ZBP1 promotes cell death in response to both Z–nucleic acids and heat stress could provide insight into how the host combats invading pathogens. Programmed cell death is an important strategy for the host to eliminate intracellular pathogens, such as viruses (*13*, *21*, *26*). By passively releasing pro-inflammatory damage-associated molecular patterns, dead cells can alert the immune system to augment inflammatory responses (*5*, *27*–*29*). Considering that infections can cause high fever, it is conceivable that hyperthermia might promote pathogen clearance and inflammation by activating ZBP1. However, extreme hyperthermia resulting from persistent environmental heat exposure could cause excessive activation of ZBP1-dependent cell death, leading to circulatory failure, DIC, multiple organ dysfunction, and even death.

REFERENCES AND NOTES

1. Y. Epstein, R. Yanovich, *N. Engl. J. Med.* **380**, 2449–2459 (2019).
2. J. F. Bobb, Z. Obermeyer, Y. Wang, F. Dominici, *JAMA* **312**, 2659–2667 (2014).
3. M. Levi, H. t. Cate, *N. Engl. J. Med.* **341**, 586–592 (1999).
4. N. Kourtis, V. Nikoletopoulou, N. Tavernarakis, *Nature* **490**, 213–218 (2012).
5. D. Wallach, T. B. Kang, C. P. Dillon, D. R. Green, *Science* **352**, aaf2154 (2016).
6. L. Sun *et al.*, *Cell* **148**, 213–227 (2012).
7. X. Chen *et al.*, *Cell Res.* **24**, 105–121 (2014).
8. N. Kayagaki *et al.*, *Nature* **526**, 666–671 (2015).
9. J. Shi *et al.*, *Nature* **526**, 660–665 (2015).
10. X. Yang *et al.*, *Immunity* **51**, 983–996.e6 (2019).
11. C. Wu *et al.*, *Immunity* **50**, 1401–1411.e4 (2019).
12. H. Jiao *et al.*, *Nature* **580**, 391–395 (2020).
13. T. Zhang *et al.*, *Cell* **180**, 1115–1129.e13 (2020).
14. R. Wang *et al.*, *Nature* **580**, 386–390 (2020).

15. K. Newton *et al.*, *Science* **343**, 1357–1360 (2014).
16. Q. Zhao *et al.*, *Cell Rep.* **19**, 798–808 (2017).
17. Y. Liu *et al.*, *Cell Death Differ.* **24**, 1459–1469 (2017).
18. H. I. Muendlein *et al.*, *Science* **367**, 1379–1384 (2020).
19. K. Newton *et al.*, *Nature* **540**, 129–133 (2016).
20. J. Lin *et al.*, *Nature* **540**, 124–128 (2016).
21. T. Kuriakose, T. D. Kanneganti, *Trends Immunol.* **39**, 123–134 (2018).
22. R. Gomez-Pastor, E. T. Burchfiel, D. J. Thiele, *Nat. Rev. Mol. Cell Biol.* **19**, 4–19 (2018).
23. A. V. Jacobsen *et al.*, *Cell Death Dis.* **7**, e2051 (2016).
24. X. M. Zhao *et al.*, *Cell Death Dis.* **7**, e2089 (2016).
25. M. Devos *et al.*, *J. Exp. Med.* **217**, e20191913 (2020).
26. J. W. Upton, W. J. Kaiser, E. S. Mocarski, *Cell Host Microbe* **11**, 290–297 (2012).
27. M. T. Lotze *et al.*, *Immunol. Rev.* **220**, 60–81 (2007).
28. M. Zheng, R. Karki, P. Vogel, T. D. Kanneganti, *Cell* **181**, 674–687.e13 (2020).
29. N. Kayagaki *et al.*, *Nature* **591**, 131–136 (2021).

ACKNOWLEDGMENTS

The authors thank J. Han, H. Zhou, and X. Xiao for sharing mouse strains. We thank L. Li, X. Y. Wang, Z. Z. Yan, and Y. T. Liu for managing mouse colonies and technical assistance. F.Y. was supported by a Postdoctoral Research Station of the 3rd Xiangya Hospital. **Funding:** This work was supported by the National Outstanding Youth Science Fund Project of the National Natural Science Foundation (no. 82025021 to B.L.), the National Natural Science Foundation of China (no. 81930059 to B.L., no. 81801888 to F.Y., and no. 81971893 to Y.T.), a key scientific project of Hunan Province (no. 2022SK2056 to B.L.), and the China Postdoctoral Science Foundation (no. 2019M652811 to F.Y.). **Author contributions:** B.L. conceived the project, supervised the research, and wrote the manuscript; B.L. and F.Y. designed the experiments; F.Y., J.C., J.W., K.Z., F.Li, F.Lia., and X.Y. performed the experiments; T.R.B., H.W., Y.T., Z.H., and L.S. commented on and edited the manuscript; and F.Y. and J.C. analyzed the data and made the figures. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials. Correspondence and requests for materials should be addressed to B.L.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abg5251
Materials and Methods
Figs. S1 to S10
MDAR Reproducibility Checklist

Submitted 12 January 2021; resubmitted 6 August 2021
Accepted 1 April 2022
10.1126/science.abg5251

REPORTS

CATALYSIS

Highly efficient catalytic production of oximes from ketones using in situ–generated H_2O_2

Richard J. Lewis[1]*, Kenji Ueura[2], Xi Liu[3,4]*, Yukimasa Fukuta[2], Thomas E. Davies[1], David J. Morgan[1,5], Liwei Chen[3,6], Jizhen Qi[7], James Singleton[1], Jennifer. K. Edwards[1], Simon J. Freakley[8], Christopher J. Kiely[9], Yasushi Yamamoto[2], Graham J. Hutchings[1]*

The ammoximation of cyclohexanone using preformed hydrogen peroxide (H_2O_2) is currently applied commercially to produce cyclohexanone oxime, an important feedstock in nylon-6 production. We demonstrate that by using supported gold-palladium (AuPd) alloyed nanoparticles in conjunction with a titanium silicate-1 (TS-1) catalyst, H_2O_2 can be generated in situ as needed, producing cyclohexanone oxime with >95% selectivity, comparable to the current industrial route. The ammoximation of several additional simple ketones is also demonstrated. Our approach eliminates the need to transport and store highly concentrated, stabilized H_2O_2, potentially achieving substantial environmental and economic savings. This approach could form the basis of an alternative route to numerous chemical transformations that are currently dependent on a combination of preformed H_2O_2 and TS-1, while allowing for considerable process intensification.

Cyclohexanone oxime is a key precursor in the production of caprolactam, a commodity chemical used in the synthesis of the polyamide nylon-6. With global production of nylon-6 predicted to reach 8.9 million metric tons per annum by 2024 (*1*), a concurrent increase in demand for cyclohexanone oxime is expected. The traditional route to cyclohexanone oxime production involves the reaction of cyclohexanone with hydroxylamine sulfate, producing ammonium sulfate—a low-value fertilizer with limited applications—as a major by-product (*2*). Alternative routes (fig. S1 and accompanying text) are hampered by the need to continually maintain a low reaction pH (by complex and energy-intensive extraction steps) or by low selectivity toward the desired product. A single-step ammoximation process that overcomes these challenges has been developed with titanium silicate-1 (TS-1) as the catalyst and cyclohexanone, ammonia, and preformed H_2O_2 as reactants (*3*). In this process, hydroxylamine is formed catalytically in situ by TS-1 (*4*), with this intermediate species subsequently reacting noncatalytically with cyclohexanone to produce the oxime (fig. S1 and accompanying text) (*5*). The catalytic activity of TS-1 with H_2O_2, which has been crucial in the development of numerous selective oxidation processes (fig. S2), is often attributed to the ability of Ti^{IV} sites to readily coordinate multiple species which, in the case of cyclohexanone ammoximation, are crucial in the formation of hydroxylamine (*6*). Despite extensive advances in catalyst design leading to the development of a range of titanosilicates that are highly selective toward cyclohexanone ammoximation and that offer greater catalytic stability [including Ti-MOR (*7*), Ti-Beta (*8*), TS-2 (*9*), and Ti-MWW (*10*)], TS-1 is still widely considered the industrial standard for reactions involving H_2O_2 (*11*).

Although the industrial ammoximation process based on H_2O_2/TS-1—which accounts for ~6 million metric tons per annum of global oxime production (*12*)—offers excellent catalytic selectivity, an excess of H_2O_2 is typically required as a result of the low stability of the oxidant under the associated reaction conditions (elevated temperatures and high pH), leading to elevated process costs (*13*). In addition, the preformed H_2O_2 requires transportation from a centralized point of production, where it is manufactured at concentrations greatly exceeding those needed in the ammoximation process; the requisite dilution effectively wastes the energy previously used in distillation and concentration steps. Furthermore, the instability of H_2O_2 necessitates the addition of acid and halide stabilizing agents to prevent its degradation during transport and storage, which in turn can limit catalyst stability, decrease reactor lifetime through corrosion, and generate substantial costs associated with their removal from product streams (*14*). Likewise, all chemical transformations that use preformed H_2O_2 suffer from these drawbacks to a certain degree.

We have previously developed catalysts for the direct synthesis of H_2O_2 from the elements that offer high synthesis rates and >99% H_2 utilization (*15*, *16*). However, to date, the direct method has been unable to rival the current

[1]Max Planck–Cardiff Centre on the Fundamentals of Heterogeneous Catalysis FUNCAT, Cardiff Catalysis Institute, School of Chemistry, Cardiff University, Main Building, Park Place, Cardiff CF10 3AT, UK. [2]UBE Corporation, 1978-5, Kogushi, Ube, Yamaguchi 755-8633, Japan. [3]School of Chemistry and Chemical, In-situ Centre for Physical Sciences, Shanghai Jiao Tong University, Shanghai 200240, P.R. China. [4]SynCat@Beijing, Synfuels China Technology Co. Ltd., Beijing 101407, P.R. China. [5]Harwell XPS, Research Complex at Harwell (RCaH), Didcot OX11 0FA, UK. [6]School of Chemistry and Chemical Engineering, Frontiers Science Centre for Transformative Molecules, Shanghai Jiao Tong University, Shanghai 200240, P.R. China. [7]i-Lab, CAS Centre for Excellence in Nanoscience, Suzhou Institute of Nano-Tech and Nano-Bionics, Chinese Academy of Sciences, Suzhou 215123, P.R. China. [8]Department of Chemistry, University of Bath, Claverton Down, Bath BA2 7AY, UK. [9]Department of Materials Science and Engineering, Lehigh University, Bethlehem, PA 18015, USA.
***Corresponding author. Email: lewisr27@cardiff.ac.uk (R.J.L.); liuxi@sjtu.edu.cn (X.L.); hutch@cardiff.ac.uk (G.J.H.)**

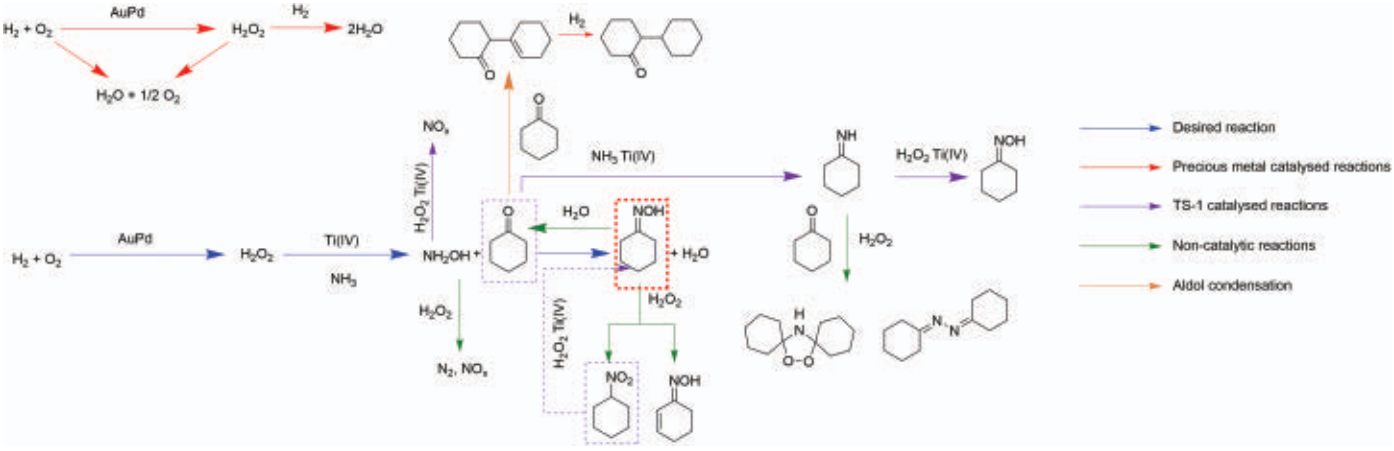


Fig. 1. Proposed key reaction pathways in the ammoximation of cyclohexanone-to-cyclohexanone oxime through in situ H_2O_2 synthesis.

industrial route to H_2O_2 production, primarily because the dilute H_2 and O_2 streams necessary to avoid potential risks of explosion limit the attainable product concentrations of H_2O_2. This disadvantage no longer applies if the H_2O_2 is produced and then rapidly consumed in situ. Indeed, the application of in situ generated H_2O_2 has been a long-standing goal in the valorization of many chemical feedstocks, with investigations into a range of selective oxidation reactions reported, including propene epoxidation (*17*), alcohol oxidation (*18*), and the partial oxidation of methane (*19*). However, earlier works have not demonstrated a viable alternative to the current respective industrial processes and are typically hindered by low rates of conversion or poor selectivity toward desired products. Indeed, in many cases the in situ approach has frequently led to unforeseen complications such as the generation of potentially hazardous by-products, often largely driven by competing hydrogenation reactions (*20*). Additionally, issues associated with catalyst deactivation and the requirement for unfavorable solvent systems or costly additives has hampered the adoption of an in situ–generated H_2O_2 route to selective oxidation (*17*).

At first sight, the application of such an in situ approach to cyclohexanone ammoximation is not only hampered by the potential drawbacks outlined in earlier studies but also by the substantial gap in conditions between the two key processes of the reaction sequence. Direct H_2O_2 formation is favored at low pH and subambient temperatures, whereas both the high reaction temperature and basic conditions associated with the ammoximation process are detrimental to H_2O_2 stability. We report that it is possible to bridge this gap in conditions and pair the direct synthesis of H_2O_2 with cyclohexanone ammoximation to produce the oxime in yields comparable to those observed in the current commercial process that uses preformed H_2O_2 (Fig. 1 and accompanying text within the supplementary materials). This is achieved through the in situ generation of H_2O_2 over AuPd nanoparticles in conjunction with a commercial TS-1 catalyst.

The immobilization of chloride-based Au and Pd salts onto a range of support materials through a wet coimpregnation procedure and subsequent calcination has been extensively reported to produce alloyed nanoparticle catalysts that are highly active toward the direct synthesis of H_2O_2 (*21*). Through this industrially viable route to catalyst preparation, we initially prepared a series of AuPd catalysts, with a range of total metal loading and supported on TiO_2, which we exposed to an oxidative heat treatment [denoted AuPd/TiO_2(Chloride-O)]. We observed a correlation between total metal loading and catalytic performance toward the direct synthesis of H_2O_2 under reaction conditions optimal for H_2O_2 production (i.e., low pH and subambient temperatures; fig. S3).

We subsequently established the high efficacy of the in situ approach to cyclohexanone ammoximation using a physical mixture of the AuPd/TiO_2(Chloride-O) catalysts of varied total metal loading in conjunction with commercial TS-1 (fig. S4 and table S1). With an optimal catalyst formulation of 0.33%Au-0.33%Pd/TiO_2(Chloride-O) (metal loading reported as weight percent), we observed an oxime yield and selectivity based on H_2 (i.e., mols of H_2 consumed that lead to the formation of the oxime through H_2O_2) of 77 and 71%, respectively. We did not observe the formation of unwanted organic by-products, such as nitrocyclohexane or cyclohexenylcyclohexanone (analysis detection threshold for by-products equivalent to ~0.005 M). Detailed characterization of the TS-1 material is presented in fig. S5, with further characterization through scanning transmission electron microscopy (STEM) presented in fig. S6, A to D. Analysis of the 0.33%Au-0.33%Pd/TiO_2(Chloride-O) catalyst by high-angle annular dark field scanning transmission electron microscopy (HAADF-STEM) imaging and x-ray energy dispersive spectroscopy (XEDS) mapping (fig. S7) revealed a bimodal particle size distribution and distinctive particle size–composition relationship: Smaller particles (3 to 10 nm) were found to be Pd-rich alloys, whereas the larger particles (10 to 30 nm) were found to be Au-rich, often adopting an Au-rich core with a Pd-rich shell morphology. Similar observations have previously been reported for AuPd catalysts prepared by this wet coimpregnation synthesis route (*22*).

Substantial improvements in cyclohexanone oxime yield were obtained by using a gaseous reactant mixture consisting of H_2 and O_2 with an N_2 diluent, in contrast to that observed when using either component alone (fig. S8 and accompanying text). Indeed, the in situ approach also offers increased cyclohexanone oxime yields (77%) compared with that observed when using preformed H_2O_2 (41%), at concentrations of H_2O_2 comparable to those that could be present if all the H_2 in the in situ reaction was converted to H_2O_2. The relatively limited activity observed when using commercial H_2O_2 can be attributed to the complete addition of H_2O_2 at the start of the reaction; continual incremental addition of H_2O_2 over the course of the reaction is well known to influence the catalytic performance of the current industrial process (fig. S8 and accompanying text). Further investigation demonstrated that high catalytic performance (selectivity toward cyclohexanone oxime >95%) could be achieved regardless of the support (TiO_2, SiO_2, CeO_2, Al_2O_3, Nb_2O_5, or ZrO_2) used to immobilize the AuPd nanoparticles (fig. S9).

Enhanced cyclohexanone oxime production was observed when both Au and Pd were immobilized onto the same support, with

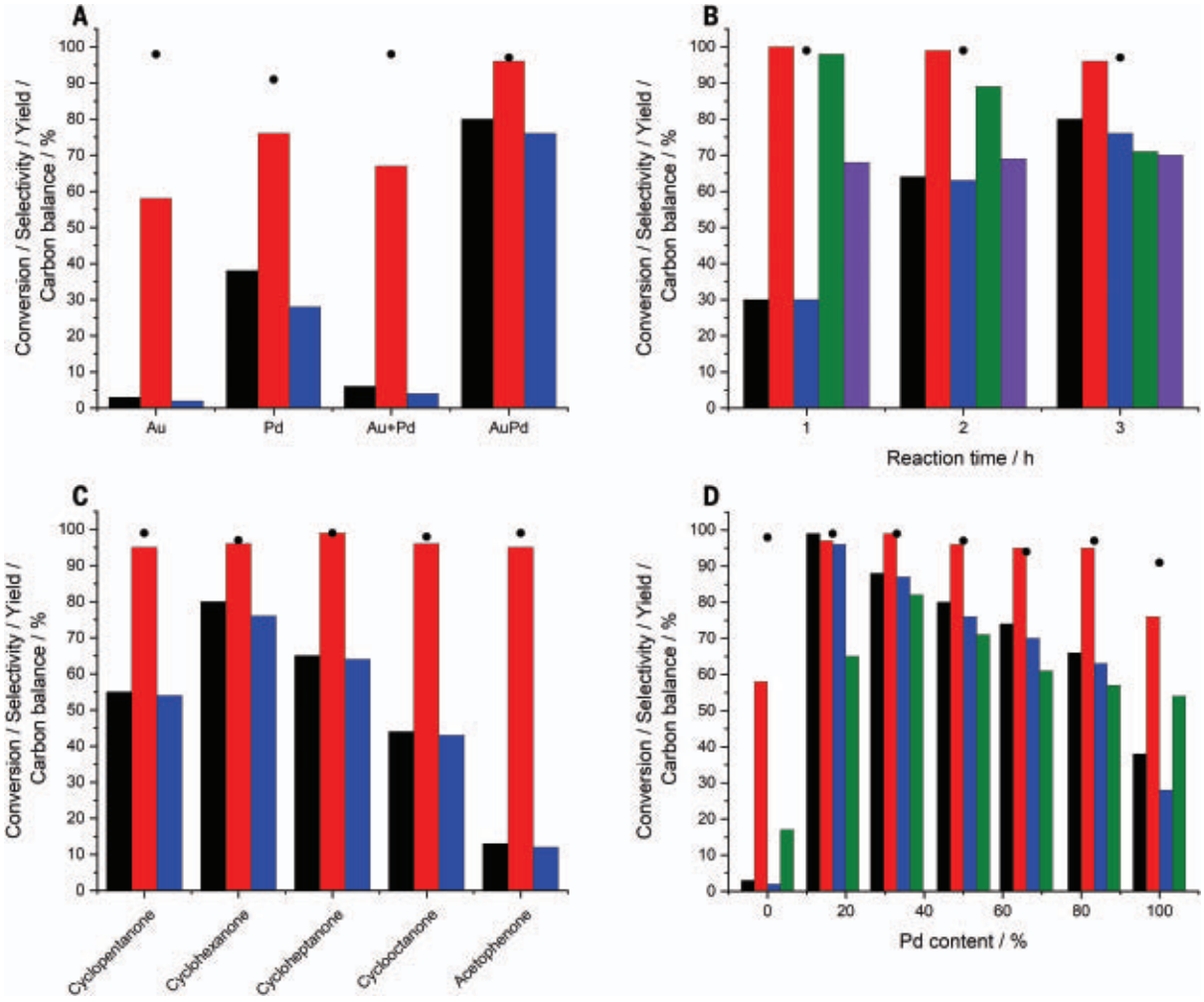


Fig. 2. Catalytic activity of supported 0.66%AuPd/TiO_2(Chloride-O) catalysts, used in conjunction with TS-1, toward the ammoximation of cyclohexanone through the in situ production of H_2O_2. (**A**) The synergistic effect of alloying Au and Pd. (**B**) Time-on-line activity of the 0.33%Au-0.33%Pd/TiO_2(Chloride-O) catalyst. (**C**) Catalytic activity of the 0.33%Au-0.33%Pd/TiO_2(Chloride-O) catalyst toward the ammoximation of a range of ketones. (**D**) The effect of the Au:Pd ratio on catalytic activity of 0.66%AuPd/TiO_2(Chloride-O) toward cyclohexanone ammoximation. Ammoximation reaction conditions: Ketone (2 mmol), NH_4HCO_3 (4 mmol), 5% H_2/N_2 (420 psi), 25% O_2/N_2 (160 psi), 0.66%AuPd/TiO_2(Chloride-O) catalyst (0.075 g), TS-1 (0.075 g), t-BuOH (5.9 g), H_2O (7.5 g), reaction time 3 hours, reaction temperature 80°C, stirring speed 800 rotations per minute (rpm). Key: Ketone conversion (black bar), selectivity toward oxime (red bar), oxime yield (blue bar), selectivity based on H_2 (green bar), selectivity based on NH_3 (purple bar), carbon balance (black circles). Note for Fig. 2A: Au refers to 0.66%Au/TiO_2 (Chloride-O), Pd to 0.66%Pd/TiO_2(Chloride-O), Au+Pd to a physical mixture of the two monometallic catalysts, and AuPd to 0.33%Au-0.33%Pd/TiO_2(Chloride-O).

the activity of the 0.33%Au-0.33%Pd/TiO_2 (Chloride-O) and TS-1 system (77% oxime yield) markedly outperforming analogs consisting of either monometallic catalysts or a physical mixture thereof (Fig. 2A and fig. S10). Indeed, the alloying of Au with Pd is known to be highly effective in both suppressing O–O bond dissociation (inhibiting H_2O production) and promoting the release of H_2O_2 from catalytic surfaces (*23*, *24*). It is therefore plausible to consider that the role of Au is to facilitate the desorption of H_2O_2 (or peroxy species) from the precious metal surface, enabling subsequent diffusion to Ti^{IV} sites present within the TS-1 framework. TS-1 in turn catalyzes the formation of the hydroxylamine intermediate, with limited ammoximation activity observed in the absence of either the titanosilicate or AuPd-supported catalyst (≤15% oxime selectivity; fig. S11). Further analyses of the 0.33% Au-0.33%Pd/TiO_2(Chloride-O) and TS-1 dual catalyst system demonstrated that high H_2 selectivity can be achieved when the reaction is not limited by cyclohexanone availability, with an H_2 selectivity of 98% observed at a reaction time of 1 hour and a cyclohexanone conversion of 30% (Fig. 2B).

We have also extended our studies to assess the ammoximation of a small range of other ketones (cyclopentanone, cycloheptanone, cyclooctanone, and acetophenone), with many of the corresponding oximes finding application in the synthesis of pharmaceuticals (*25*, *26*). Using the 0.33%Au-0.33%Pd/TiO_2 (Chloride-O) and TS-1 catalysts (Fig. 2C), oxime selectivities >95% were observed for all substrates, demonstrating the versatility of the in situ approach to oxime formation. The variation in the rate of ketone conversion is primarily related to the differing intrinsic reactivity of the ketones with hydroxylamine, in addition to the limited ability of the larger ketones to access the interior of the titanosilicate pore structure (*27*).

Further optimization of the Au:Pd ratio revealed an optimal composition with a physical mixture of the 0.55%Au-0.11%Pd/TiO_2(Chloride-O) and TS-1 catalysts exhibiting a cyclohexanone oxime yield of 96% and an apparent turnover frequency (TOF) (223 $mol_{oxime}\ mol_{metal}^{-1} hour^{-1}$), far greater than that of the physical mixture of the 0.33% Au-0.33%Pd/TiO_2(Chloride-O) catalyst and TS-1 (143 $mol_{oxime} mol_{metal}^{-1} hour^{-1}$) in a 3-hour

reaction (Fig. 2D, apparent TOFs shown in table S2). The improved catalytic performance may be due in part to a decrease in the number of contiguous Pd sites, which are known to be highly active toward O-O bond cleavage (*28*) and enhanced H_2O_2 utilization. Our determination of H_2 selectivity supports this hypothesis, with the combination of Au-rich catalysts in addition to TS-1 offering enhanced selectivity in comparison to the corresponding Au-lean analogs (table S2). In a similar manner, Au-rich compositions are observed to offer reduced rates of H_2O_2 degradation under ideal direct synthesis conditions (fig. S12).

Although high yields of cyclohexanone oxime can be achieved using the cocatalyst system, application on an industrial scale would require the use of a composite catalyst that can both synthesize H_2O_2 and catalyze the formation of hydroxylamine. Hence, we investigated the efficacy of AuPd nanoparticles supported on a commercial TS-1 for the ammoximation of cyclohexanone, with the catalyst prepared by wet coimpregnation of $PdCl_2$ and $HAuCl_4$ onto TS-1 followed by calcination [denoted 0.33%Au-0.33%Pd/TS-1(Chloride-O)]. The rate of H_2O_2 formation by this catalyst was comparable to that observed when using a range of oxide supports (fig. S13) [characterization of 0.33%Au-0.33%Pd/TS-1(Chloride-O) is presented in fig. S14], with the bifunctional catalyst offering >95% selectivity toward the oxime (fig. S15). However, the rate of cyclohexanone conversion (44%) was lower than that observed for the cocatalyst system, which constituted the 0.33%Au-0.33%Pd/TiO_2(Chloride-O) catalyst in conjunction with TS-1 (80%). This difference in catalytic performance is attributable to the blocking of Ti^{IV} sites, as previously observed by Hölderich and coworkers (*29*), and to the poor mixing of the Au and Pd metallic components upon immobilization onto the titanosilicate support as evidenced by STEM-HAADF imaging and XEDS mapping (fig. S16).

We subsequently demonstrated that it is possible to further enhance catalytic performance through preparation of the titanosilicate-supported catalyst through a sequential wet impregnation procedure using $Pd(OAc)_2$ and $HAuCl_4$ precursors, followed by calcination [denoted 0.33%Au-0.33%Pd/TS-1(Acetate-O)]

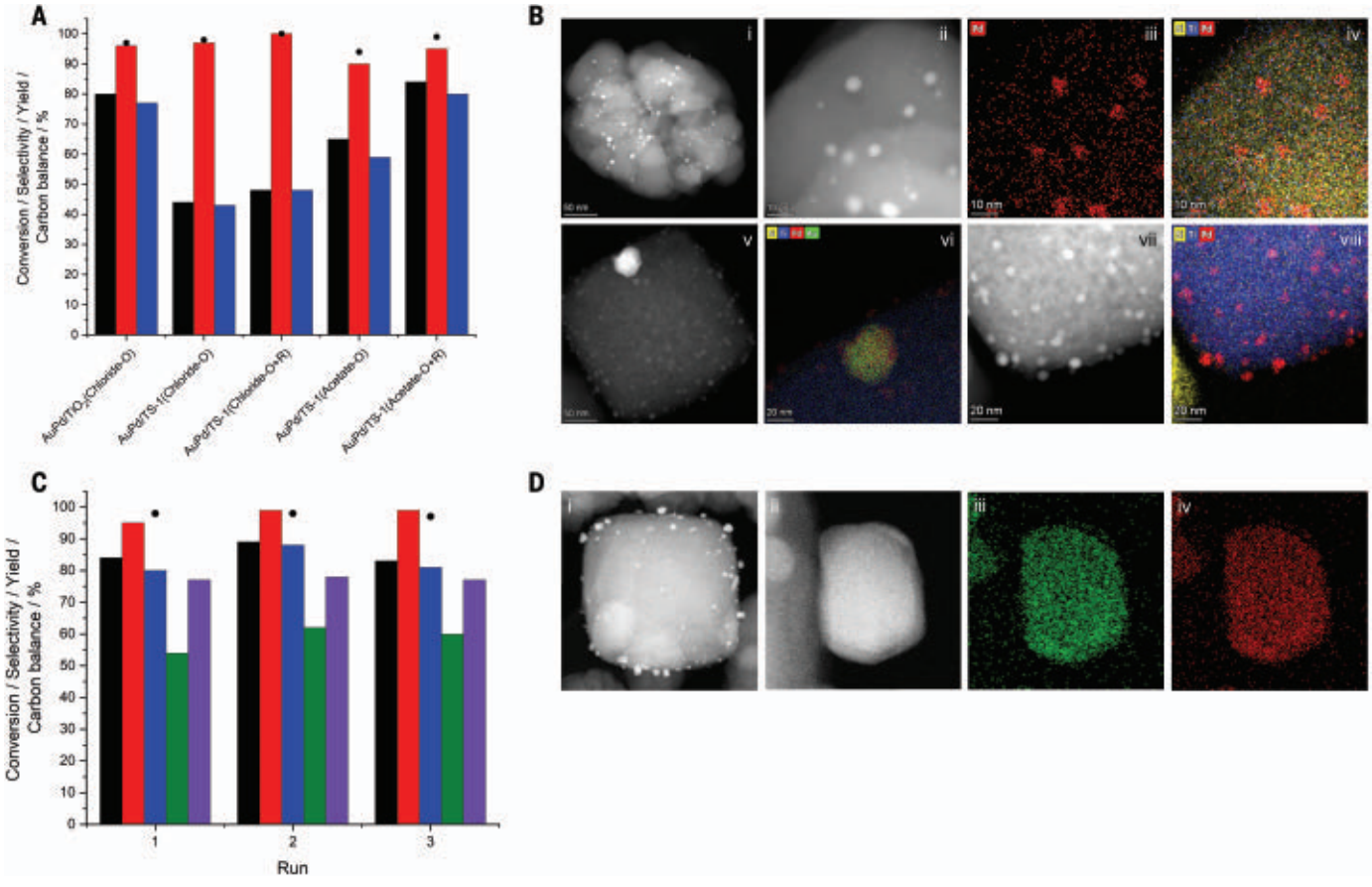


Fig. 3. Performance and stability of composite AuPd catalysts supported on TS-1 toward the ammoximation of cyclohexanone through in situ production of H_2O_2. (**A**) Catalytic activity as a function of the Pd precursor and heat treatment regime. (**B**) Microstructural analysis of the unused 0.33%Au-0.33%Pd/TS-1(Acetate-O+R) sample, which had been calcined (flowing air, 400°C, 3 hours) and then reduced (5%H_2/Ar, 400°C, 2 hours). (i) Lower and (ii) higher magnification HAADF-STEM images of the titanosilicate majority component. (iii) XEDS map showing Pd metal attachment only and absence of Au on TS-1. (iv) XEDS overlay map showing Si (yellow), Pd (red), and low-concentration Ti (blue). (v) HAADF-STEM image of the TiO_2 minority component showing considerable metal attachment with (vi) corresponding XEDS elemental mapping [Si (yellow), Ti (blue), Pd (red), and Au (green)] showing that the larger metal particles on the TiO_2 particles are AuPd alloys. (vii and viii) HAADF-STEM and corresponding XEDS maps of the smaller Pd-only particles, respectively. (**C**) Catalytic reusability of the 0.33% Au-0.33%Pd/TS-1(Acetate-O+R) catalyst in a batch regime. (**D**) Microstructural analysis of the 0.33%Au-0.33%Pd/TS-1(Acetate-O+R) catalyst after three consecutive ammoximation reactions. Low (i) and high (ii) magnification HAADF-STEM images of the TiO_2 minority component and STEM-XEDS elemental mapping of Au (green) (iii) and Pd (red) (iv), showing stability and retention of the AuPd alloy nanoparticles post reaction. Ammoximation reaction conditions: Cyclohexanone (2 mmol), NH_4HCO_3 (4 mmol), 5%H_2/N_2 (420 psi), 25%O_2/N_2 (160 psi), catalyst (0.075 g), t-BuOH (5.9 g), H_2O (7.5 g), reaction time 3 hours, reaction temperature 80°C, stirring speed 800 rpm. Key: Cyclohexanone conversion (black bar), selectivity toward oxime (red bar), oxime yield (blue bar), carbon balance (black circles). Note for Fig. 3A: 0.33%Au-0.33%Pd/TiO_2 (Chloride-O) used in conjunction with TS-1 (0.075 g), all other conditions as stated.

and a subsequent reductive heat treatment (2 hours, 400°C, 5%H_2/Ar) [denoted 0.33% Au-0.33%Pd/TS-1(Acetate-O+R) (characterization of the 0.33%Au-0.33%Pd/TS-1(Acetate-O+R) catalyst is reported in fig. S17)]. Catalytic performance toward cyclohexanone ammoximation was found to be markedly improved compared with the 0.33%Au-0.33%Pd/TS-1(Chloride-O) or 0.33%Au-0.33%Pd/TS-1(Chloride-O+R) analogs, with the yield of oxime achieved by the 0.33%Au-0.33%Pd/TS-1(Acetate-O+R) catalyst comparable to that observed when using the 0.33%Au-0.33%Pd/TiO_2(Chloride-O) and TS-1 physical mixture (Fig. 3A, with comparison of apparent TOFs shown in table S3). Indeed, our optimal results using in situ synthesized H_2O_2 rival those reported in the literature for a range of commonly used oxidants, including preformed H_2O_2 (table S4), demonstrating the potential of this approach to supersede the current industrial route to cyclohexanone oxime.

With the nature of the catalyst surface—particularly the oxidation state of the active metals crucial in obtaining high catalytic performance—we analyzed the titanosilicate-supported AuPd catalysts by x-ray photoelectron spectroscopy (XPS) (fig. S18). Exposure of the 0.33%Au-0.33%Pd/TS-1(Acetate-O) catalyst to a reductive heat treatment (2 hours, 400°C, 5%H_2/Ar) resulted in a complete shift in the Pd oxidation state to Pd^0, coinciding with an observed increase in catalytic performance toward both H_2O_2 synthesis (fig. S19) and cyclohexanone ammoximation (Fig. 3A). Detailed STEM analysis of the 0.33%Au-0.33%Pd/TS-1(Acetate-O+R) catalyst (Fig. 3B; additional analysis shown in fig. S20) identified significant metal decoration on both the TiO_2 minority and TS-1 majority phases. HAADF-STEM imaging and corresponding XEDS elemental mapping revealed that the metal nanoparticles present on the minority TiO_2 phase consisted predominantly of larger (5 to 20 nm) AuPd alloys, in addition to some smaller (1 to 3 nm) Pd-only particles. Comparative analysis of the TS-1 majority phase shows the preferential immobilization of Pd, with the notable absence of Au or AuPd alloys.

Time-on-line studies conducted using the 0.33%Au-0.33%Pd/TS-1(Acetate-O+R) catalyst showed that high H_2 selectivity (94%) is possible when cyclohexanone availability is not limited, indicating that at extended reaction times H_2 is nonselectively consumed through H_2O_2 degradation or noncatalytic pathways (fig. S21), whereas ammonia selectivity, regardless of reaction time, was found to be relatively high (~75%).

Analysis of postreaction solutions by inductively coupled plasma mass spectrometry (table S5) revealed the high stability of Au over a standard 3-hour cyclohexanone ammoximation reaction. However, a substantial loss of Pd (18.6%) was observed over this same time period. Further studies established the stability of the 0.33%Au-0.33%Pd/TS-1(Acetate-O+R) catalyst over multiple uses and found that there was some minimal additional leaching of Pd upon second use (table S5). Notably, no metal loss was observed after the third use, and indeed the efficacy of the catalyst was retained over three consecutive ammoximation reactions (oxime yield ≥80%) (Fig. 3C).

Detailed STEM-HAADF analysis of the catalyst over three uses (Fig. 3D and figs. S22 to S24) identified that the observed Pd leaching was associated with the loss of the smaller nonalloyed Pd nanoparticles, which were found predominantly on the TS-1 majority phase in the as-prepared material. The composition and dispersion of the AuPd nanoalloys present on the minority TiO_2 component were retained after multiple uses. Hot filtration experiments in which the 0.33%Au-0.33%Pd/TS-1(Acetate-O+R) catalyst was replaced by bare TS-1 revealed that there was no contribution of leached species toward the formation of cyclohexanone oxime (fig. S25, A and B, and accompanying text). However, in the absence of the immobilized precious metals (i.e., when only bare TS-1 was present) some additional conversion of the ketone was observed (7%), which is in keeping with our previous observations (fig. S11) and can be attributed to the ability of TS-1 to promote the formation of unwanted by-products in the absence of H_2O_2 (*30*). Further studies, using reagent concentrations much greater than those used during the catalytic studies and comparable to those used under the industrial process, revealed the increased stability of the 0.33%Au-0.33%Pd/TS-1(Acetate-O+R) catalyst compared with a monometallic Pd analog, and indicated that a combination of NH_3 and H_2O_2 is responsible for promoting the dissolution of active metals (table S6). In keeping with our STEM-HAADF analysis (Fig. 3D and figs. S22 to S24), the alloying of Au with Pd was found to substantially inhibit metal leaching, even under harsh reaction conditions. These observations, when coupled with our earlier studies comparing the activity of supported AuPd catalysts with monometallic analogs (Fig. 2A) and the negligible activity of homogeneous Pd species (fig. S25, A and B, and accompanying text), suggest that the AuPd alloy nanoparticles supported on the TiO_2 minority component of the TS-1 support are key to achieving high catalytic performance, whereas the unalloyed Pd nanoparticles are largely spectator species.

A major challenge of the current industrial route to cyclohexanone oxime is associated with the deactivation of the TS-1 catalyst, through formation of TiO_2-SiO_2 domains, induced by the presence of relatively high concentrations of ammonia in reactant streams (*31*). In keeping with these observations, our analysis by XPS (fig. S26 and table S7) indicated a minor shift in Ti speciation over sequential reuse in the ammoximation reaction, indicative of the formation of TiO_2-like surface species, although no loss in catalyst activity was observed. In recent years hollow titanium silicates (HTS-1) have been developed through the postsynthesis treatment of TS-1 (*32*), with these materials found to offer far greater stability when used in the ammoximation reaction than the parent material (*33*). As such, we consider that upon potential industrial application, any deactivation of the TS-1 component can be readily overcome through the adoption of HTS-1 or alternative titanosilicate support.

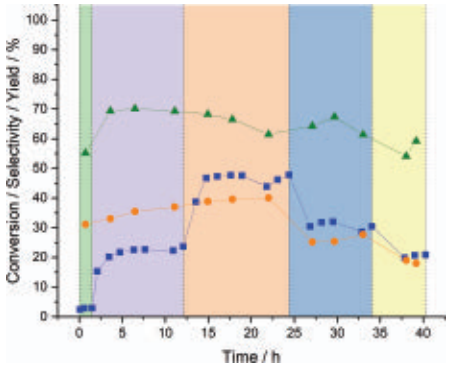


Fig. 4. Optimization of reaction parameters for the 0.33%Au-0.33%Pd/TS-1(Acetate-O+R) catalyst in a continuous regime. Ammoximation reaction conditions: Cyclohexanone (20 wt %): NH_3 (26 wt %) (1: 0.5), 3.6% H_2, 6.4% O_2, 90% N_2 (580 psi, 20 ml min^{-1}), catalyst (0.41 g), 0.33%Au-0.33%Pd/TS-1(Acetate-O+R): Al_2O_3 (4:1) t-BuOH: H_2O (9:1, 0.01 to 0.10 ml min^{-1}), residence time 76 min at 0.01ml min^{-1} liquid flow rate, reaction temperature 80°C. Key: Cyclohexanone oxime yield (blue squares), H_2 conversion (orange circles), H_2 selectivity (green triangles). Reaction conditions between 0 and 1.5 hours: as above with liquid flow of 0.1 ml min^{-1} (green background). Reaction conditions between 1.5 and 12.1 hours: as above with liquid flow of 0.02 ml min^{-1} (purple background). Reaction conditions between 12.1 and 24.4 hours: as above with liquid flow of 0.01 ml min^{-1} (orange background). Reaction conditions between 24.4 and 34.0 hours: as above with liquid flow of 0.01 ml min^{-1} and total pressure of 290 psi (blue background). Reaction conditions between 34.0 and 41.4 hours: as above with liquid flow of 0.01 ml min^{-1} and total pressure of 145 psi (yellow background). Note: Given the ratio of cyclohexanone: NH_3 used in this study (1:0.5) it is possible to conclude that under optimal reaction conditions (orange background), NH_3 selectivity approaches 100% (96%), given the near 50% (48%) oxime yield observed and the stoichiometry of the ammoximation reaction.

We have previously demonstrated that supported AuPd nanoparticles offer high stability toward H_2O_2 production in a flow regime under conditions optimized for H_2O_2 selectivity (*34*). With these observations in mind and with the independence of the in situ route to the ammonia source established (table S8), we next evaluated the stability of the 0.33%Au-0.33%Pd/TS-1(Acetate-O+R) catalyst toward the ammoximation of cyclohexanone via in situ H_2O_2 production, through use of a continuous flow reactor (fig. S27) and concentrations of cyclohexanone and ammonia comparable to those used in the industrial ammoximation process. In this case, the catalyst was exposed to an oxidative heat treatment (16 hours, 110°C, static air) before reduction (2 hours, 200°C, H_2) and exhibited comparable performance to analogous materials exposed to higher-temperature heat treatments (Fig. 3A) under batch conditions (fig. S28). These continuous flow studies demonstrated that the high catalytic stability observed under batch conditions (Fig. 3C) can be readily translated to a flow system (Fig. 4) in which liquid [cyclohexanone and ammonia (NH_3 aq.)] and gaseous (H_2 and O_2) reagents are continuously introduced into the reactor. Indeed, using this continuous flow reactor, cyclohexanone oxime yield and H_2 selectivity were observed to be steady over several hours on stream at 48 and 70%, respectively, with no observable loss in catalytic stability detected over 40 hours of reaction. Moreover, given the limited availability of ammonia [cyclohexanone: NH_3 (aq.) 1:0.5], a near-complete selective utilization of this reagent was observed (96% ammonia selectivity), which would avoid the substantial costs associated with reagent separation and recycling upon any potential industrial application of the in situ process. Further investigation revealed that the cyclohexanone oxime yield can be increased considerably through reaction condition optimization (87%) (fig. S29), although in keeping with our earlier studies (Fig. 2B), H_2 selectivity was found to be inherently linked to cyclohexanone availability.

In an attempt to establish the industrial viability of the in situ route to cyclohexanone ammoximation and with a focus on the 0.33%Au-0.33%Pd/TS-1(Acetate-O+R) catalyst, we conducted extended lifetime studies under industrially relevant reaction conditions and over 248 hours on stream (fig. S30; flow reactor schematic shown in fig. S31). Although analysis of the postreaction catalyst through x-ray diffraction did not indicate any substantial loss in TS-1 crystallinity (fig. S32, i to ii), our XPS evaluation revealed a slight shift in Ti speciation (fig. S32, iii to iv), possibly indicative of the formation of TiO_2 domains within the titanosilicate component as previously observed during industrial application (*30*). Our analysis by STEM revealed no significant agglomeration of precious metal nanoparticles (fig. S32, v to viii) and the maintenance of the AuPd nanoalloys over the course of the reaction. This is in keeping with the high stability of the AuPd species previously observed under batch conditions (figs. S22 to S24) and further highlights the long-term stability of the catalyst under prospective industrial conditions.

Finally, we conducted a detailed technoeconomic evaluation in which we compared the in situ approach and the current industrial process, which uses preformed H_2O_2, on the basis of previous evaluations made by Zhu *et al.* (*35*) (fig. S33, on the basis of flow data provided in Fig. 4, with further detail provided in table S9). Assuming that the lifetime of the 0.33%Au-0.33%Pd/TS-1(Acetate-O+R) catalyst is comparable to that reported for TS-1 in the current industrial route and a comparable activity of the TS-1 component (i.e., 0.3 kg of TS-1 catalyzes 1 ton cyclohexanone oxime production) (*36*), our calculations demonstrate the economic viability of the in situ approach, and we estimate savings of 13% based on material costs alone (assuming a catalyst lifetime of 2.3 years). Indeed, even assuming a far more limited catalyst lifetime (0.75 years) our economic evaluation reveals that the material cost of the in situ approach is comparable to that of the current industrial process. This evaluation does not account for the substantial savings associated with in situ H_2O_2 production, namely those associated with transport and storage of H_2O_2 and increased reactor longevity, as reactor corrosion is known to result from the presence of the stabilizing agents in preformed H_2O_2 (*37*). Additionally, considerable environmental savings are associated with a less carbon-intensive manufacturing process of this key platform chemical. As such, the in situ route represents a positive step toward more sustainable selective chemical transformations and has the potential to supersede the current industrial route to cyclohexanone oxime. More broadly, we consider that this approach may find wider application in other industrial oxidation reactions that at present are dependent on the use of TS-1 with preformed H_2O_2.

REFERENCES AND NOTES

1. HDIN Research, Global Nylon 6 production capacity to reach 8.86 million tons in 2024 (2019); www.hdinresearch.com/news/56
2. R. Mokaya, M. Poliakoff, *Nature* **437**, 1243–1244 (2005).
3. P. Roffia *et al.*, *Stud. Surf. Sci. Catal.* **55**, 43–52 (1990).
4. P. Roffia *et al.*, (Montedipe SpA) US Patent US4794198A (1987).
5. A. Thangaraj, S. Sivasanker, P. Ratnasamy, *J. Catal.* **131**, 394–400 (1991).
6. A. Zecchina *et al.*, *Catal. Today* **32**, 97–106 (1996).
7. Y. Xu *et al.*, *Chem. Eng. Sci.* **152**, 717–723 (2016).
8. L. Xu, J. Ding, Y. Yang, P. Wu, *J. Catal.* **309**, 1–10 (2014).
9. J. S. Reddy, S. Sivasanker, P. Ratnasamy, *J. Mol. Catal.* **69**, 383–392 (1991).
10. S. Zhao *et al.*, *Appl. Catal. A*, **394**, 1–8 (2011).
11. Sumitomo to commercialize new caprolactam process; https://www.chemicalonline.com/doc/sumitomo-to-commercialize-new-caprolactam-pro-0001.
12. B. Zong *et al.*, *Engineering* **3**, 379–384 (2017).
13. A. C. K. Yip, X. Hu, *Ind. Eng. Chem. Res.* **50**, 13703–13710 (2011).
14. R. J. Lewis, G. J. Hutchings, *ChemCatChem* **11**, 298–308 (2019).
15. J. K. Edwards, B. Solsona, *Science* **323**, 1037–1041 (2009).
16. S. J. Freakley *et al.*, *Science* **351**, 965–968 (2016).
17. Q. Chen, E. J. Beckman, *Green Chem.* **10**, 934–938 (2008).
18. C. M. Crombie *et al.*, *ACS Catal.* **11**, 2701–2714 (2021).
19. Z. Jin *et al.*, *Science* **367**, 193–197 (2020).
20. G. Wang *et al.*, *Chem. Catalysis* **1**, 885–895 (2021).
21. R. J. Lewis *et al.*, *ChemCatChem* **11**, 1673–1680 (2019).
22. G. J. Hutchings, C. J. Kiely, *Acc. Chem. Res.* **46**, 1759–1772 (2013).
23. J. Li, T. Ishihara, K. Yoshizawa, *J. Phys. Chem. C* **115**, 25359–25367 (2011).
24. T. Richards *et al.*, *Nat. Catal.* **4**, 575–585 (2021).
25. A. B. Fernandez, I. Lezcano-Gonzalez, M. Boronat, T. Blasco, A. Corma, *Phys. Chem. Chem. Phys.* **11**, 5134–5141 (2009).
26. L. Li, M. Chen, F. C. Jiang, *Bioorg. Med. Chem.* **24**, 1853–1865 (2016).
27. P. Wu, T. Komatsu, T. Yashima, *J. Catal.* **168**, 400–411 (1997).
28. T. Ricciardulli *et al.*, *J. Am. Chem. Soc.* **143**, 5445–5464 (2021).
29. W. Laufer, R. Meiers, W. Hölderich, *J. Mol. Catal. Chem.* **141**, 215–221 (1999).
30. A. Cesana, M. A. Mantegazza, M. Pastori, *J. Mol. Catal. Chem.* **117**, 367–373 (1997).
31. G. Petrini *et al.*, *Stud. Surf. Sci. Catal.* **68**, 761–766 (1991).
32. Y. Wang, M. Lin, A. Tuel, *Microporous Mesoporous Mater.* **102**, 80–85 (2007).
33. C. Xia *et al.*, "Environmental-friendly catalytic oxidation processes based on hierarchical titanium silicate zeolites at SINOPEC" in *Green Chemical Processing and Synthesis*, I. Karame, H. Srour, Eds. (Intech Open, 2017), chap 6, pp. 119–150.
34. S. J. Freakley *et al.*, *ACS Catal.* **3**, 487–501 (2013).
35. Z. Zhu *et al.*, *Energy Convers. Manage.* **192**, 100–113 (2019).
36. U. Romano, M. Ricci, "Industrial applications" in *Liquid Phase Oxidation via Heterogeneous Catalysis: Organic Synthesis and Industrial Applications*, M. G. Clerici, O. A. Kholdeeva, Eds. (Wiley, 2013), p. 470.
37. G. Gao *et al.*, *Chin. J. Catal.* **41**, 1039–1047 (2020).
38. R. J. Lewis *et al.*, doi: 10.17035/d.2022.0163568304.

ACKNOWLEDGMENTS

We appreciate technical support from H. Matsumoto and C. Zeng, Hitachi High-Technologies (Shanghai) Co., Ltd. for HR-STEM characterization, and the Cardiff University electron microscope facility for the transmission electron microscopy. **Funding:** The authors thank UBE Corporation for financial support. XPS data collection was performed at the EPSRC National Facility for XPS ("HarwellXPS"), operated by Cardiff University and UCL, under contract PR16195. R.J.L. and G.J.H. gratefully acknowledge Cardiff University and the Max Planck Centre for Fundamental Heterogeneous Catalysis (FUNCAT) for financial support. X.L. acknowledges financial support from the National Natural Science Foundation of China (22872163 and 22072090). L.C. acknowledges financial support from the National Natural Science Foundation of China (21991153). In addition, S.J.F. acknowledges the award of a Prize Research Fellowship from the University of Bath. **Author contributions:** R.J.L., K.U., Y.F., S.J.F., and G.J.H. contributed to the design of the study; R.J.L. and K.U. conducted experiments and data analysis. R.J.L., K.U., Y.F., J.S., J.K.E., S.J.F., C.J.K., Y.Y. and G.J.H. provided technical advice and result interpretation. R.J.L., X.L., T.E.D., D.J.M., L.C., J.Q., and C.J.K. conducted catalyst characterization and corresponding data processing. R.J.L., C.J.K., and G.J.H. wrote the manuscript; R.J.L. and C.J.K. wrote the supplementary materials; all authors commented on and amended both documents. All authors discussed and contributed to the work. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The data supporting the findings of this study are available within the article and its supplementary materials or from the authors upon reasonable request, with the underlying data found at the Cardiff University Data Repository (*38*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abl4822
Materials and Methods
Supplementary Text
Figs. S1 to S33
Tables S1 to S9
References (*39*–*80*)

Submitted 19 July 2021; resubmitted 12 October 2021
Accepted 30 March 2022
10.1126/science.abl4822

QUANTUM OPTICS

Tunable and state-preserving frequency conversion of single photons in hydrogen

R. Tyumenev[1]†, J. Hammer[1], N. Y. Joly[2,1]*, P. St. J. Russell[1], D. Novoa[3,4,1]

In modern quantum technologies, preservation of the photon statistics of quantum optical states upon frequency conversion holds the key to the viable implementation of quantum networks, which often require interfacing of several subsystems operating in widely different spectral regions. Most current approaches offer only very small frequency shifts and limited tunability, while suffering from high insertion loss and Raman noise originating in the materials used. We introduce a route to quantum-correlation–preserving frequency conversion using hydrogen-filled antiresonant-reflecting photonic crystal fibers. Transient optical phonons generated by stimulated Raman scattering enable selective frequency up-conversion by 125 terahertz of the idler photon of an entangled pair, with efficiencies up to 70%. This threshold-less molecular modulation process preserves quantum correlations, making it ideal for applications in quantum information.

Light quanta are efficient and reliable carriers of information because they barely interact with the environment. In this regard, the manipulation of quantum states of light such as single photons or entangled photon pairs without affecting their quantum properties has attracted major interest in the last two decades (*1*, *2*). Whereas operations on spatial and temporal photonic degrees of freedom can be carried out using passive devices (*3*, *4*), other modifications such as frequency conversion must be performed in active systems via nonlinear interactions (*5*–*13*). Quantum-correlation–preserving frequency conversion is essential in modern quantum technologies requiring quantum light sources operating at different frequencies, but is quite difficult to achieve. First, the quality of the quantum source is degraded by almost unavoidable addition of noise from Raman scattering or fluorescence during interactions with the media within which the light propagates. Second, apart from the special case of thin, high-gain media (*14*), most approaches based on second- or third-order nonlinear processes require stringent phase-matching conditions to be satisfied, which in the common case of dispersive solid materials effectively restricts the spectral range of operation to a few selected frequencies and greatly narrows the bandwidth over which the conversion can efficiently take place.

Gas-filled antiresonant-reflecting photonic crystal fiber (ARR-PCF) allows highly efficient nonlinear frequency conversion by a combination of pressure-adjustable dispersion and nonlinearity, with diffraction-free low-loss broadband guidance (*15*–*17*). When filled with a Raman-active diatomic gas such as hydrogen, transient optical phonons or coherence waves of synchronous collective molecular motion can be excited in the gaseous core via stimulated Raman scattering (SRS). They can be subsequently used for threshold-less frequency up-conversion of arbitrary signals via molecular modulation (*18*, *19*), provided momentum conservation is satisfied (*20*).

Here we report the use of Raman coherence waves excited in H_2-filled ARR-PCFs for efficient quantum frequency conversion of a single photon from a highly correlated photon pair (a biphoton; Fig. 1A). As the up-conversion process depends strongly on phase matching between the interacting single photons and coherence waves, the signal is almost free from any background noise caused by incoherent molecular modulation. Spontaneous Raman scattering from the solid microstructure is also negligible because the light-glass overlap is extremely low (< 0.01%) (*17*, *21*). In addition, other parasitic nonlinear effects such as instantaneous four-wave mixing are also largely suppressed in the gaseous core owing to the low effective nonlinearity at the pump powers used.

In SRS-based quantum molecular modulation, a strong pump pulse launched in the fundamental core mode (Fig. 1B) is scattered by hydrogen molecules in the core, giving rise to a Stokes signal down-shifted by the Raman frequency ν_R ~125 THz [i.e., the frequency of the main vibrational transition of the *Q*-branch of H_2 (*22*)]. The pump-Stokes beat note then drives the molecular vibrations, stimulating exponential amplification of the Stokes light and causing excitation of a quantum coherence wave of collective molecular vibrations. The result (Fig. 1A) is a moving refractive index grating capable of threshold-less Doppler frequency shifting of any subsequent, arbitrarily weak signal, providing phase matching is satisfied, i.e., dephasing $\vartheta = |(\beta_{ui} - \beta_i) - \beta_{cw}| \approx 0$ (*20*), where β_i, β_{ui}, and β_{cw} are the propagation constants of the idler photon, its up-shifted counterpart, and the coherence wave, respectively. In this regard, the "*S*-shaped" dispersion landscape of gas-filled ARR-PCFs (*20*) (Fig. 1C) allows selective intramodal phase-matched conversion of any pair of frequencies from the ultraviolet to the mid-infrared—a working range not achievable by any other existing technology—provided they spectrally differ by ν_R. In our system, $\vartheta \approx 0$ occurs at ~70 bar (Fig. 1C).

To verify these predictions experimentally, we split the output of a laser delivering 1064-nm pulses of 3.8-ns duration into two arms. A small fraction was used to generate the entangled biphotons at 849 nm (signal) and 1425 nm (idler) through spontaneous four-wave mixing (*23*, *24*) in a suspended-core fiber (fig. S2). The signal photons were directly sent to a superconducting single-photon detector (SSPD) and the idler photons, along with the remaining pump power, were launched into a 60-cm-long, single-ring–type ARR-PCF with 60-μm core diameter and ~270-nm average capillary-wall thickness (Fig. 1B). These design parameters ensured that loss-inducing anticrossings between the core mode and resonances in the glass-walled capillaries (*21*) lay spectrally away from any wavelength of interest. The launch efficiencies for the pump in the fundamental mode (Fig. 1B) reached 96%, which helped to mitigate noise arising from unwanted effects in the glass microstructure. Both fiber ends were enclosed in hermetically sealed gas cells. The out-coupled idler photons, whether frequency up-converted to 894 nm or not, were sent to a second SSPD, and coincidences between the two channels were recorded using a time tagger (*24*), enabling access to the second-order correlation function $g^{(2)}(\tau)$, where $\tau = t_i - t_s$ is the difference between the idler (t_i) and signal (t_s) arrival times. The set-up is sketched in fig. S3.

We first measured the quantum conversion efficiency of the up-shifting process (Fig. 2A), defined as $\eta = [1 - (I_{RC}/I)] \times 100$, where I_{RC} and I are the idler count rates with and without Raman coherence prepared in the system (controlled by the presence of 115-μJ pump pulses; Fig. 2B). This is a valid measure of conversion efficiency in an effective two-level system such as the hydrogen molecule because no other viable scattering routes (such as rotational SRS (*12*) or vibrational down-shifting) are available to the idler photons. As shown in Fig. 2A, η reaches 70% at 70 bar, in excellent agreement with the theory, and gradually falls at other pressures owing to strong

[1]Max-Planck Institute for the Science of Light, Staudtstrasse 2, 91058 Erlangen, Germany. [2]Department of Physics, Friedrich-Alexander-Universität, Staudtstrasse 2, 91058 Erlangen, Germany. [3]Department of Communications Engineering, Engineering School of Bilbao, University of the Basque Country (UPV/EHU), Torres Quevedo 1, 48013 Bilbao, Spain. [4]IKERBASQUE, Basque Foundation for Science, Plaza Euskadi 5, 48009 Bilbao, Spain.
*Corresponding author. Email: nicolas.joly@mpl.mpg.de
†Present address: Interherence GmbH, Henkestrasse 91, 91052 Erlangen, Germany.

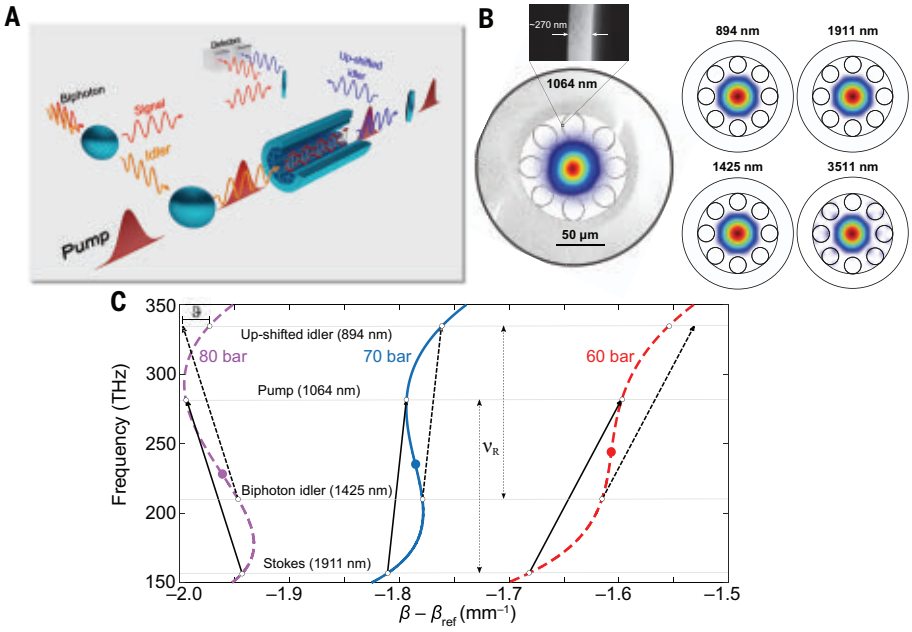


Fig. 1. Fiber-based Raman molecular modulation of quantum light sources. (**A**) Illustration of the nonlinear conversion process of single idler photons from biphotons. (**B**) Near-field mode profile of the pump (1064 nm) imaged by a charge-coupled device camera and overlaid with a scanning electron micrograph of the transverse fiber microstructure. (Right) theoretical intensity profiles (linear scale) of fundamental modes at the wavelengths of the Stokes signal (1911 nm), the biphoton idler (1425 nm), and its up-shifted anti-Stokes (894 nm) and down-shifted Stokes (3511 nm) bands, obtained via finite-element modeling of a perfect structure. (**C**) Frequency versus wave vector diagrams at three different pressures. To reveal the subtle *S*-shaped profile, the modal propagation constants are offset by $\beta_{ref} = (2\pi\nu/c)n_{eff}(\nu_{ref})$, where n_{eff} is the modal index at ν_{ref} = 600 THz and ν is the optical frequency. The arrows represent the coherence waves (solid) and the corresponding single-photon transitions (dashed), which perfectly match at ~70 bar and are otherwise highly dephased. The filled-in circles mark the zero-dispersion points.

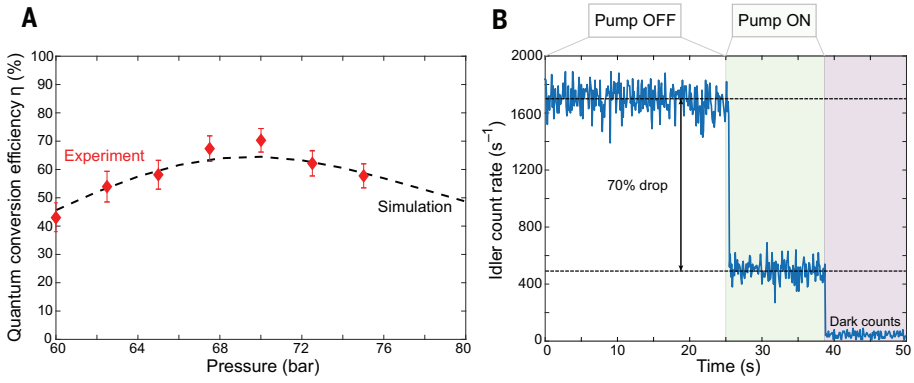


Fig. 2. Pressure dependence of the quantum conversion efficiency. (**A**) Conversion efficiency as a function of gas pressure. The experimental results (symbols) are in good agreement with numerical simulations (dashed line) based on Maxwell-Bloch equations. (**B**) Estimate of the conversion efficiency based on the drop in idler count rate at 1425 nm at 70 bar. When the pump signal is unblocked after 25 s of measurement time (shaded green area), the excited optical phonons yield a clear drop in the idler count rate (see fig. S1 for the behavior of the up-shifted idler). Only dark counts are registered (shaded purple area) when all the optical signals are blocked.

dephasing and Raman gain suppression (*25*), the latter being responsible for the total cancellation of the modulation process below ~50 bar (fig. S4).

The excellent quantitative agreement between experimental results and numerical modeling (Fig. 2A) (*24*) allows us to gain insight into the quantum molecular modulation process in the fiber. In particular, the spatiotemporal evolution of the Raman coherence can be studied, something not accessible in the experiments. As illustrated in Fig. 3A, at 70 bar the Raman coherence builds up along the fiber, peaking at ~30 cm. From this point on, the single idler photons are predominantly transformed into anti-Stokes idler photons at 894 nm (Fig. 3B), with conversion to Stokes photons remaining five orders of magnitude lower (Fig. 3C). In the last section of the fiber, the pump pulses are substantially depleted and the coherence gradually decays away (Fig. 3A). Despite the high pressures, backward SRS does not play any role in the dynamics (fig. S5).

It is also important to verify that the preexisting quantum correlations in the original biphoton pairs are preserved during single-pump molecular modulation and subsequent propagation of the up-shifted anti-Stokes photons through the ARR-PCF. To do so, we measured the second-order correlation function $g^{(2)}(\tau)$ of the original biphotons (Fig. 4A) and compared the result with the second-order correlations established between the up-shifted idler and the unmodified signal (Fig. 4B). The strong resemblance of both $g^{(2)}(\tau)$ peaks (Fig. 4C), normalized to account for accidental coincidences caused by the pulsed nature of the biphoton source (*24*, *26*), as shown in fig. S6, indicates that fragile quantum correlations remain intact during the Raman molecular modulation process. It is important to highlight that the system is, in principle, fully reconfigurable by using other Raman-active molecular degrees of freedom or different gas species and mixtures (*25*). Furthermore, because state-of-the-art ARR-PCFs feature extraordinarily low attenuation (*27*) and polarization-maintaining properties (*28*, *29*), these fiber-based solutions might also provide environmentally insulated transport of the modulated biphotons, making these quantum devices far more versatile than any other alternative to date.

We have demonstrated that quantum molecular modulation based on SRS in gas-filled ARR-PCFs can be used for efficient, quantum-correlation–preserving frequency conversion of biphoton pairs. The approach is independent of the quality and nature of the quantum

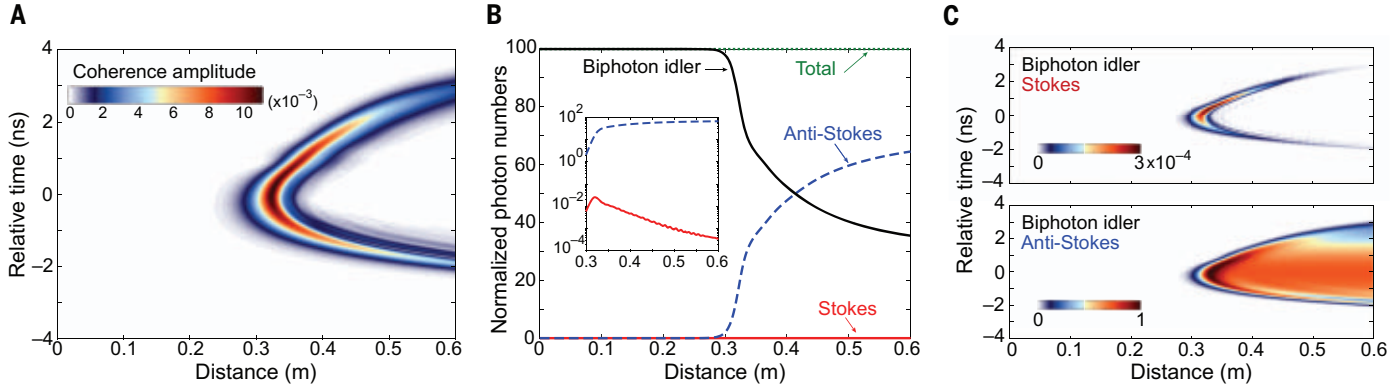


Fig. 3. Numerical modeling of the frequency shifting process at 70 bar. (**A**) Spatiotemporal evolution of Raman coherence inside the fiber. Time is measured relative to a reference frame moving at the group velocity of the pump pulses. (**B**) Evolution of the photon numbers normalized so that the efficiency η can be directly extracted. The signals displayed are the original idler (black solid line), its Stokes (red solid line) and anti-Stokes (blue-dashed line) bands, and their total sum (green-dotted line). The inset shows the detailed evolution over the last 30 cm of fiber length (logarithmic vertical scale). (**C**) Spatiotemporal evolution of the Stokes and anti-Stokes idler bands, normalized to the peak of the anti-Stokes. The Stokes signal peaks after 30 cm and weakens as it propagates further, owing to strong dephasing. In sharp contrast, the anti-Stokes light is continuously amplified up to the fiber end face.

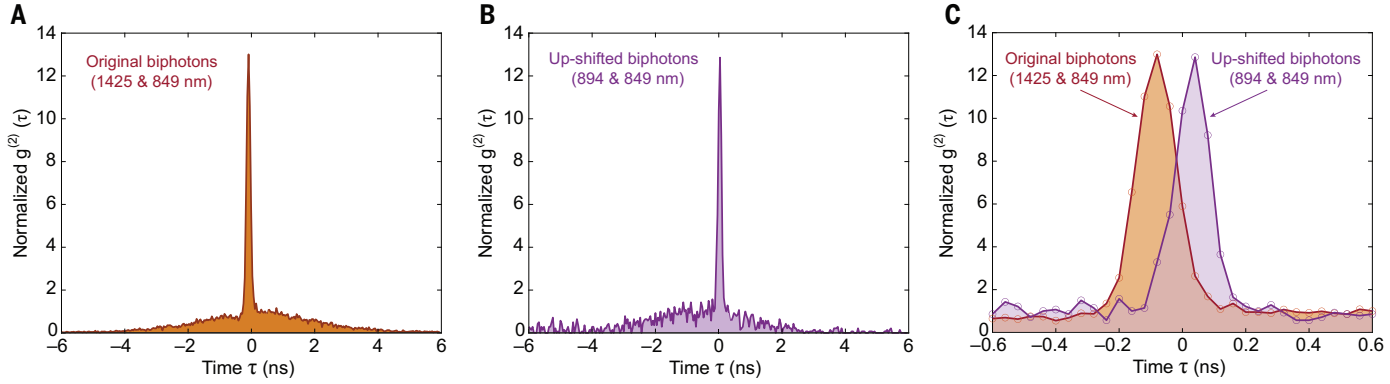


Fig. 4. Unaffected quantum correlations. (**A**) Second-order correlations of the original biphoton pairs and (**B**) of the upshifted-idler + original signal biphotons. The $g^{(2)}(\tau)$ peaks sit on a pedestal of accidental coincidences. The ratio of both $g^{(2)}(\tau)$ peaks is ~0.99 and the difference of their widths is ~25 ps, well below the minimum time step of ~50 ps resolvable with our set-up. (**C**) Close-up of both measurements. The correlation peaks are noncoincident in time owing to group-velocity dispersion in the fibers delivering the single photons to the SSPDs.

source used, meaning that it will preserve any degree of second-order correlations present in the original system. Moreover, the excellent mode quality inherent to the system may facilitate its efficient interconnection with other quantum devices (*2*). These advantages, along with the unparalleled performance of ARR-PCFs, provides a route for multiple technological applications in fields such as quantum-enhanced imaging (*30*) and communications (*1*).

REFERENCES AND NOTES

1. H. J. Kimble, *Nature* **453**, 1023–1030 (2008).
2. D. Awschalom *et al*., *PRX Quantum* **2**, 017002 (2021).
3. J. B. Spring *et al*., *Science* **339**, 798–801 (2013).
4. K. Y. Bliokh, F. J. Rodríguez-Fortuño, F. Nori, A. V. Zayats, *Nat. Photonics* **9**, 796–808 (2015).
5. X. Li, J. Chen, P. Voss, J. Sharping, P. Kumar, *Opt. Express* **12**, 3737–3744 (2004).
6. H. J. McGuinness, M. G. Raymer, C. J. McKinstrie, S. Radic, *Phys. Rev. Lett.* **105**, 093604 (2010).
7. K. De Greve *et al*., *Nature* **491**, 421–425 (2012).
8. A. S. Clark, S. Shahnia, M. J. Collins, C. Xiong, B. J. Eggleton, *Opt. Lett.* **38**, 947–949 (2013).
9. M. Allgaier *et al*., *Nat. Commun.* **8**, 14288 (2017).
10. L. Fan *et al*., *Nat. Photonics* **10**, 766–770 (2016).
11. N. Matsuda, *Sci. Adv.* **2**, e1501223 (2016).
12. P. J. Bustard, D. G. England, K. Heshami, C. Kupchak, B. J. Sussman, *Phys. Rev. A* **95**, 053816 (2017).
13. T. A. Wright, C. Parry, O. R. Gibson, R. J. A. Francis-Jones, P. J. Mosley, *Opt. Lett.* **45**, 4587–4590 (2020).
14. A. V. Sokolov *et al*., *Opt. Lett.* **26**, 728–730 (2001).
15. F. Benabid, J. C. Knight, G. Antonopoulos, P. S. J. Russell, *Science* **298**, 399–402 (2002).
16. F. Couny, F. Benabid, P. J. Roberts, P. S. Light, M. G. Raymer, *Science* **318**, 1118–1121 (2007).
17. P. S. J. Russell, P. Hölzer, W. Chang, A. Abdolvand, J. C. Travers, *Nat. Photonics* **8**, 278–286 (2014).
18. S. E. Harris, A. V. Sokolov, *Phys. Rev. A* **55**, R4019–R4022 (1997).
19. J. Z. Li, M. Katsuragawa, M. Suzuki, K. Hakuta, *Phys. Rev. A* **58**, R58–R60 (1998).
20. M. K. Mridha, D. Novoa, P. Hosseini, P. St. J. Russell, *Optica* **6**, 731–734 (2019).
21. C. Wei, R. J. Weiblen, C. R. Menyuk, J. Hu, *Adv. Opt. Photonics* **9**, 504–561 (2017).
22. W. K. Bischel, M. J. Dyer, *J. Opt. Soc. Am. B* **3**, 677 (1986).
23. J. Hammer, M. V. Chekhova, D. R. Häupl, R. Pennetta, N. Y. Joly, *Phys. Rev. Res.* **2**, 012079 (2020).
24. Materials and methods are available as supplementary materials.
25. P. Hosseini, D. Novoa, A. Abdolvand, P. S. J. Russell, *Phys. Rev. Lett.* **119**, 253903 (2017).
26. O. A. Ivanova, T. Sh. Iskhakov, A. N. Penin, M. V. Chekhova, *Quantum Electron.* **36**, 951–956 (2006).
27. S.-F. Gao *et al*., *Nat. Commun.* **9**, 2828 (2018).
28. A. Taranta *et al*., *Nat. Photonics* **14**, 504–510 (2020).
29. P. Roth *et al*., *Optica* **5**, 1315–1321 (2018).
30. C. A. Casacio *et al*., *Nature* **594**, 201–206 (2021).
31. R. Tyumenev, J. Hammer, N. Y. Joly, P. St. J. Russell, D. Novoa, Data for: Tunable and state-preserving frequency conversion of single-photons in hydrogen, *Zenodo* (2022); https://doi.org/10.5281/zenodo.5910082.

ACKNOWLEDGMENTS

We thank M. V. Chekhova and C. Genes for useful discussions.
Funding: Max-Planck Society (R.T., J.H., N.Y.J., P.S.J.R., D.N.).
Author contributions: Conceptualization: D.N., N.Y.J., P.S.J.R.

Funding acquisition: N.Y.J., P.S.J.R. Investigation: R.T., D.N. Methodology: R.T., J.H., N.Y.J., P.S.J.R., D.N. Software: R.T., J.H., D.N. Supervision: D.N., N.Y.J., P.S.J.R. Visualization: R.T., J.H., N.Y.J., P.S.J.R., D.N. Writing: R.T., J.H., N.Y.J., P.S.J.R., D.N.
Competing interests: The authors declare that they have no competing interests. **Data and materials availability:** All data from this study, including the numerical simulation codes, are available at Zenodo (*31*).

SUPPLEMENTARY MATERIALS
science.org/doi/10.1126/science.abn1434

Materials and Methods
Figs. S1 to S6
References (*32*–*34*)

Submitted 5 November 2021; accepted 8 February 2022
10.1126/science.abn1434

HYDROGELS

Cuticular pad–inspired selective frequency damper for nearly dynamic noise–free bioelectronics

Byeonghak Park[1], Joo Hwan Shin[1], Jehyung Ok[1], Subin Park[1], Woojin Jung[1], Chanho Jeong[2], Seunghwan Choy[1], Young Jin Jo[1], Tae-il Kim[1,2,3]*

Bioelectronics needs to continuously monitor mechanical and electrophysiological signals for patients. However, the signals always include artifacts by patients' unexpected movement (such as walking and respiration under approximately 30 hertz). The current method to remove them is a signal process that uses a bandpass filter, which may cause signal loss. We present an unconventional bandpass filter material—viscoelastic gelatin-chitosan hydrogel damper, inspired by the viscoelastic cuticular pad in a spider—to remove dynamic mechanical noise artifacts selectively. The hydrogel exhibits frequency-dependent phase transition that results in a rubbery state that damps low-frequency noise and a glassy state that transmits the desired high-frequency signals. It serves as an adaptable passfilter that enables the acquisition of high-quality signals from patients while minimizing signal process for advanced bioelectronics.

Advanced bioelectronics such as wearable (*1*, *2*) and implantable (*3*, *4*) devices are considered highly promising for the continuous detection and measurement of human physiological signals (*5*, *6*). Although recent accomplishments involving the integration of soft and flexible materials with ultrathin electronics (*7*, *8*) have enabled continuous monitoring (*9*, *10*) and multifunctionality (multimodal sensing and stimulation) (*11*–*13*), such applications (for example, in patient care) are generally limited by signal artifacts that arise from dynamic noise (usually under 30 Hz) (*14*, *15*). These signal artifacts include noise arising from patient movements, such as breathing, walking, tapping, and running. To selectively remove the dynamic noise embedded in these biosignals, processing techniques such as bandpass filtering are used (*14*). However, that can cause loss of information (*15*) and hardly change the band on demand (table S1). And wirelessly transfer of their whole signals, including noises in miniaturized wearable electronics, suffer from signal delay in the signal processing. Although damping materials that have shock-absorbing properties could be useful, they need to be designed to damp the wide or selective ranges of frequency of biophysiology.

[1]School of Chemical Engineering, Sungkyunkwan University (SKKU), Suwon 16419, Republic of Korea. [2]Department of Biomedical Engineering, SKKU, Suwon 16419, Republic of Korea. [3]Biomedical Institute for Convergence at SKKU (BICS), SKKU, Suwon 16419, Republic of Korea.
*Corresponding author. Email: taeilkim@skku.edu

Spiders can use their webs to monitor minute vibration signals generated by their prey, enemies, and mates, even in noisy (windy or rainy) conditions, which usually correspond to low-frequency (~30 Hz) vibrations (Fig. 1A) (*16*). Spiders can separate target vibration signals from mechanical noise using a selective vibration frequency damping organ in the form of a cuticular pad located under the vibration receptor (fig. S1) (*16*, *17*). The cuticular pad has viscoelastic properties that result from viscous bonds such as hydrogen bonds between chitin and protein chains, and the pad material phase changes from a rubbery to a glassy state above an applied frequency of near 30 Hz (Fig. 1B) (*17*). This phase transition allows the pad to selectively transmit target vibration signals (higher than 30 Hz) and filter low-frequency noise (lower than 30 Hz).

We fabricated a chitosan and gelatin interpenetrating hydrogel damper (fig. S2 and movie S1). The dynamic mechanical properties of the hydrogel are similar to those of the spider's cuticular pad, with modulus increasing with frequency (Fig. 1C). The material enables viscous damping below the transition frequency. Above the transition frequency, the tan δ value decreases while the modulus increases with frequency, so that vibrations elastically transmit. Attributed to its low-frequency damping, the hydrogel can be used in eliminating dynamic noises under 30 Hz (Fig. 1D and fig. S3). The key mechanism we hypothesize for selective damping is based on the relaxation time of the viscoelastic material, determining the material's transition frequency (Fig. 1E). The ratio of the relaxation time to the deformation time (observation time) is defined as *De*, which characterizes the viscoelasticity of the material. When a vibration stimulus reaches the hydrogel damper, the weak bonds responsible for the viscosity partially break and absorb the vibration energy, resulting in a rubbery state. If the relaxation time is less than the vibration period ($De < 1$; viscous) and the mechanical stress is quickly recovered before the next period of the vibration, then the vibration is continuously absorbed. As the frequency increases up to the transition point ($De = 1$), shear thickening occurs, and the viscosity increases because of the friction between the chains with remaining viscosity-related bonds, and the damping is maximized. However, if the relaxation time is longer than the vibration period ($De > 1$; elastic), meaning that the stress does not recover before the next period of the vibration, then the absorption fails, the chains rearrange, and the vibration transmits.

To investigate the selective damping property of the hydrogel damper composed of semi-interpenetrating hydrogels with gelatin and chitosan, we implemented dynamic mechanical analysis through the frequency sweep with chitosan, gelatin, and gelatin-chitosan hydrogel (Fig. 1F). The chitosan maintains a high tan δ value because of its viscous bonds between its chains in overall frequency, whereas gelatin has an apparent phase transition (*18*, *19*). We hypothesize that the external vibration stimulus dissociates the viscous bonds, such as hydrogen bonds, of the polymer matrix from dominantly chitosan hydrogel, dissociates hydrophobic interactions from dominantly gelatin hydrogel, and rearranges chitosan and gelatin chains in the manner of the stimulus (fig. S2) (*20*). Mixtures of gelatin and chitosan exhibit both high damping and phase transition, so that the selective frequency damping can be realized. The mechanical compression test shows decreasing hysteresis on the frequency (fig. S4) by shear-thickening damping properties, and the damping mechanism predominantly comes from the break and recovery of the viscous weak bonds (fig. S5). The highpass filter characteristics can be shown with 14.18 dB/decade at 27°C and 59.69 dB/decade at 45°C in a Bode plot (fig. S6 and table S2).

These selective frequency-damping properties can be used in bioelectronics. Dynamic

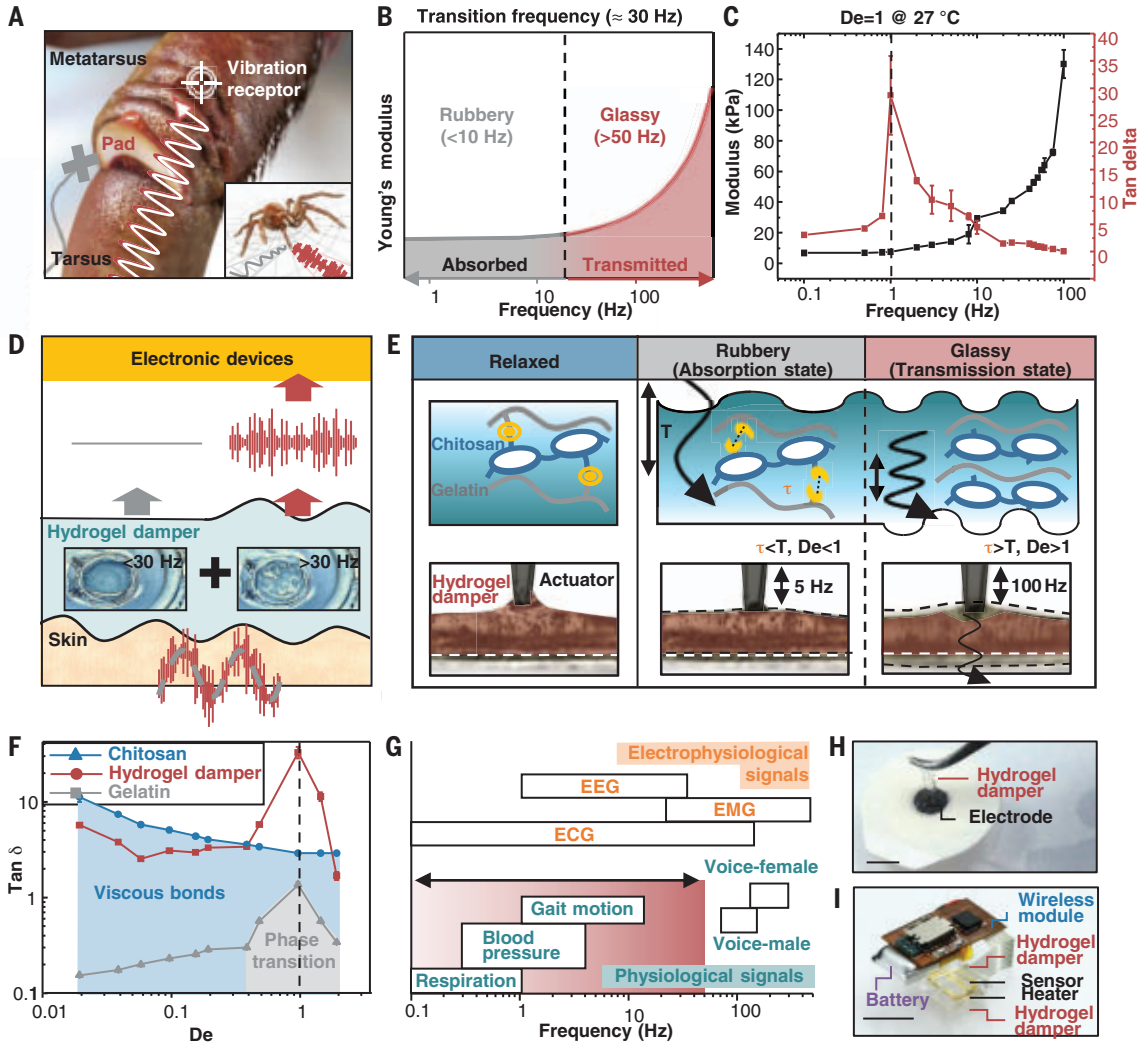


Fig. 1. Selective noise damping in a spider's cuticular pad and the bioinspired gelatin-chitosan hydrogel damper for selective frequency-dependent damping. (**A**) Schematic illustrations of selective low-frequency damping of superposed high-frequency target signals (>30 Hz) at the viscoelastic cuticular pad between the tarsus and metatarsus. The selectively transmitted target signals reach the lyriform slit organs (vibration receptor). (**B**) Viscoelastic properties of the spider's pad, which depend on the applied frequency of vibration (*28*). The mechanical properties of the damping pad include a rubber-glass transition under an ~30-Hz vibration. (**C**) Dynamic mechanical storage modulus and tan δ of the gelatin-chitosan hydrogel versus frequency (n = 3 samples, mean ± SD). (**D**) Schematic illustration of selective biophysiological signal detection through the hydrogel damper from skin and the surroundings. At high frequency, the vibration starts to transmit through the hydrogel damper surface, with the plate vibration patterns visualized on the surface. (**E**) Selective damping mechanism in the hydrogel damper related to the relaxation time. The transition between absorption and transmission is determined by the *De* of the hydrogel damper. The hydrogel damper on 1.5 μm polyethylene terephthalate (PET) (white dotted lines) fluctuates under a 100-Hz vibration, transmitting the vibrations *De* > 1. (**F**) Dynamic tan δ curves of chitosan, gelatin, and the gelatin-chitosan hydrogel below and above *De* = 1. (**G**) Representative frequency ranges of human mechanical (blue) and electrophysiological (orange) biosignals and mechanical damping ranges of the hydrogel damper (red) in the 27° to 45°C temperature range (*21*–*23*). (**H** and **I**) Example of bioelectronics by the hydrogel damper electrode for measuring electrophysiological signals (H) with mechanical noise and (I) incorporated with the hydrogel damper for selective mechanical target signal detection by a crack-based strain sensor connected to a Bluetooth module. Scale bars, (H) 1 cm; (I) 8 mm.

noise from motion artifacts is found over a wide frequency range (Fig. 1G), from 0.01 to nearly 15 Hz, including even involuntary mechanical signals, such as respiration (0.1 to 1 Hz), heartbeat (0.3 to near 4 Hz), and gait motion (1 to 15 Hz) signals (*21*–*23*). Integration with the hydrogel material on conventional electrodes can be compatible to detect electrophysiological signals such as electroencephalogram (EEG) and electrocardiogram (ECG) signals (Fig. 1H). The hydrogel damper electrode targets the elimination of low-frequency mechanical noise while transmitting electrical biophysiological signals through an electrical pathway. To apply our selective damping material to the detection of distinct mechanical biophysiological signals, we designed a prototype device that consists of a 2-mm-thick hydrogel coated on a mechanosensor. We demonstrate a nanoscale crack-based mechanosensor inspired by a spider's vibration-sensing slit organ, with a Bluetooth module for wireless data acquisition and a closed-loop compensation system for temperature variation by use of a heater and thermocouple (Fig. 1I and fig. S7).

Engineering the relaxation time of the hydrogel damper can shift the damping curve and transmitted frequencies according to users' demand. When the polymeric chains have a higher diffusion rate, they have more chances to recover their stress, leading to fast relaxation time under the given mechanical stimulus (Fig. 2A). The relaxation time is determined by the temperature and molecular weight of gelatin, which affects the diffusion rate of the chains, as confirmed with the Rouse model in tube theories (Fig. 2B and figs. S2 and S8) (*24*). The water content can be one of the factors for the viscoelasticity, but for stability, the concentration was fixed at 6 wt % on the basis of the optimization (fig. S9). The relaxation time

changes from 1.11 to less than 0.0193 s (Fig. 2B). Drop-ball tests with a 3.2-mm-diameter stainless steel ball on the hydrogel substrates represent elastic transmission and reflection of mechanical energies at 19°C, where the *De* number is ≫1, whereas the damping of mechanical energies occurs after the temperature is increased to 45°C, where *De* is ≪1 (Fig. 2C and movie S2). The viscoelastic properties can be controlled by tuning the temperature, as exemplified by our ability to shift the damping frequency from 0.89 to 51.8 Hz by changing the temperature from 27° to 45°C (Fig. 2D); in this way, the damping frequency can be tuned according to the user requirements (fig. S10).

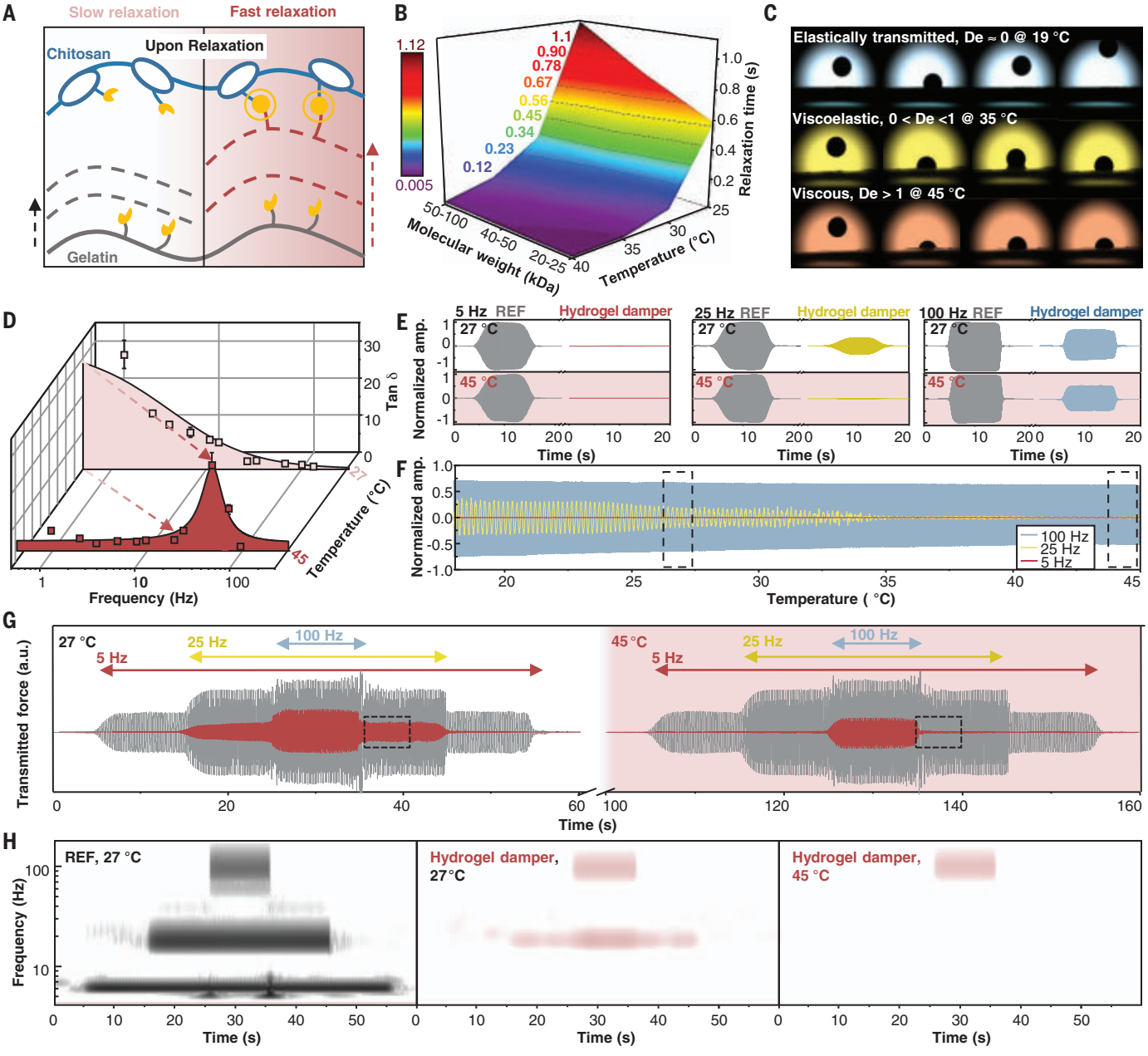


Fig. 2. Shifting the damping region by tuning the relaxation time, which corresponds to the transition frequency. (**A**) Schematic illustration of chain relaxation after a vibration stimulus in relatively (left) low-temperature and (right) high-temperature atmospheres. A higher temperature accelerates the relaxation. (**B**) Temperature and molecular weight dependency of the relaxation time and transition frequency of the hydrogel damper. The transition frequency is the reciprocal of the observed relaxation time when *De* = 1. (**C**) Viscoelastic transmission and damping at different temperatures (19°, 35°, and 45°C), with *De* verified by dropping metal balls on the hydrogel layer. (**D**) Shifting of the hydrogel damper transition frequency from 0.89 Hz at 27°C to 51.8 Hz at 45°C. (**E**) Measured real-time waveforms of vibrations in reference (gray) and the hydrogel damper samples with 5- (red), 25- (yellow), and 100-Hz (sky blue) frequencies individually applied at ambient temperature (27°C, top) and in a 45°C atmosphere (bottom). (**F**) Normalized waveforms of transmitted vibrations with 5 (red), 25 (yellow), and 100 Hz (sky blue) frequencies individually applied upon a temperature change from 18° to 45°C. The dotted box indicates the waveforms at ambient temperature (27°C) and 45°C. (**G**) Measured real-time waveform of vibrations in a reference (gray) and the hydrogel damper (red) with superposed 5-, 25-, and 100-Hz multiple-vibration frequencies applied at (left) ambient temperature (27°C) and (right) in a 45°C atmosphere. (**H**) Morlet wavelet transform of each vibration transmitted through a reference at 27°C and the hydrogel damper at 27°C and 45°C.

Transmitted vibration amplitudes through the hydrogel damper (Fig. 2E, red) and polydimethylsiloxane (PDMS) reference (Fig. 2E, gray) at 27° and 45°C with 10 s of vibrations (5, 25, and 100 Hz) and a temperature sweep (from 18° to 45°C) at three applied vibration frequencies (5, 25, and 100 Hz) (Fig. 2F) implies that selective frequency damping and shifting of the frequencies is feasible (fig. S8). For the 25-Hz vibration, there is a marked decrease in transmission with the temperature that reflects the temperature-induced transition-point shift compared with the continuous-absorption 5 Hz and transmission 100 Hz (Fig. 2F). Even with multiple superposed vibrations, selective damping by the hydrogel damper at 5 Hz occurs at 27°C (Fig. 2G), whereas raising the temperature to 45°C enables the damping frequency to be increased to 51.8 Hz. By contrast, the PDMS elastomer transmits vibrations regardless of frequency and temperature because of its elastic properties. Morlet wavelet transform images demonstrate the selective damping properties and shifts in the hydrogel damper damping curve between 27° and 45°C; the damping by the reference is shown for comparison (Fig. 2H).

We compared the damping properties with commercially available polymeric dampers—D3O (*25*), Alphagel (Theta 7) (*26*), a silicone-based elastomer (PDMS), and a shear-thickening cornstarch suspension—that are viscoelastic shear thickening materials for shock absorption. When each damper's absorbed energy per volume is measured at various frequencies to investigate energy absorption capability (Fig. 3A), the peak damping energies in the hydrogel damper case, at both 27° and 45°C, are at least 6.7 times higher (113.04 mJ/mm^3 at 27°C and 108.62 mJ/mm^3 at 45°C) than the next best-performing damping material. Even though the damping curve can be shifted by tuning the temperature, the amount of absorbed energy shows the only minimal change. Attributed to the apparent phase transition of the hydrogel damper, the relaxation time changes, and transition frequencies eventually show dramatic changes under the temperature (Fig. 3B). The absorbed energy comparison of the hydrogel damper shows superior selectivity between the noise and target signals (338.73 at 27°C and 282.39 at 45°C), which is more than 20 times those of the other damping materials (Fig. 3C). The hydrogel damper also exhibits a damping factor that is more than 3.35 times larger (22 mJ·s/m^3 at 27°C and 5.7 mJ·s/m^3 at 45°C)

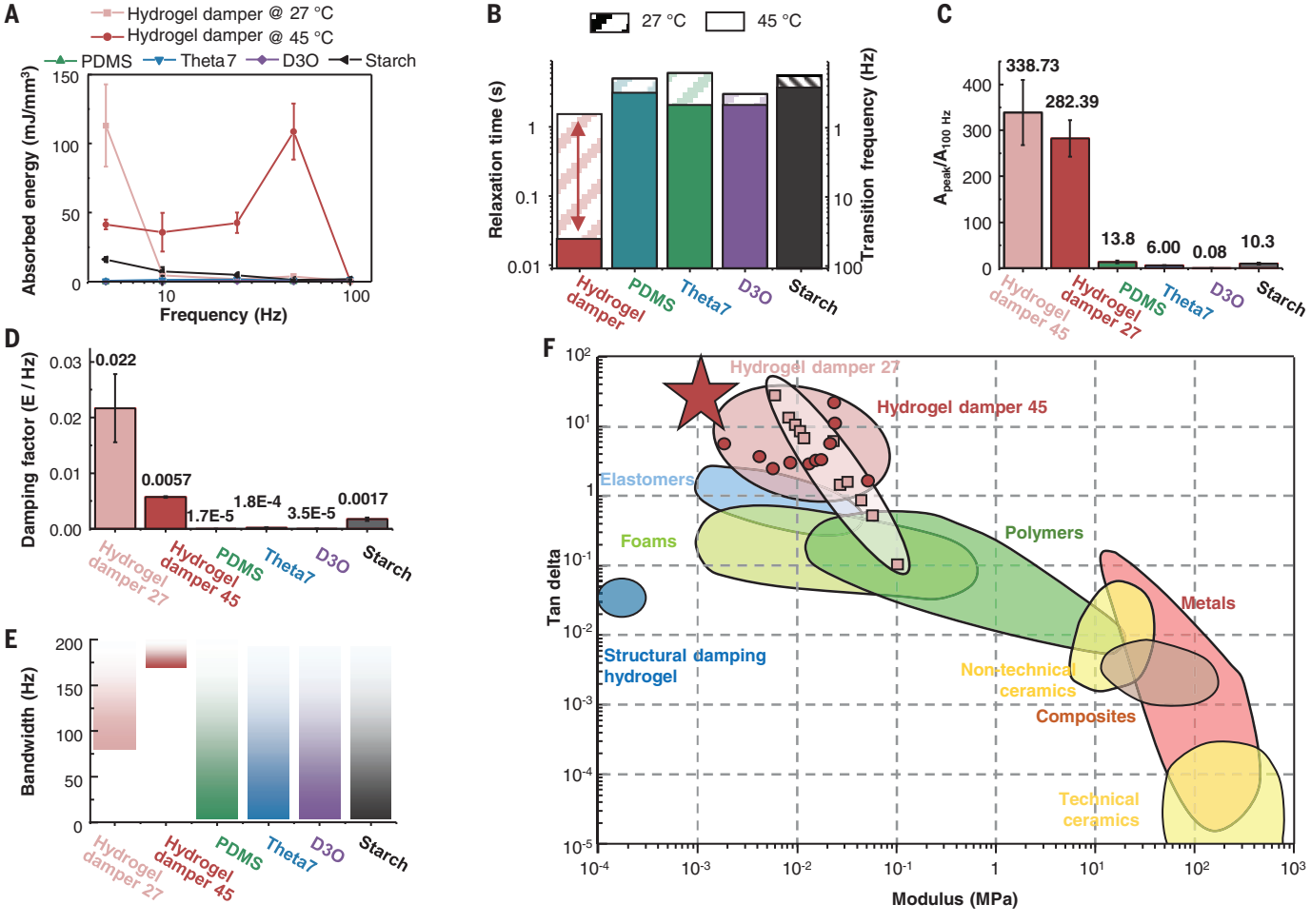


Fig. 3. Damping properties of the gelatin-chitosan hydrogel damper compared with other damping materials. (**A**) Absorbed energy per volume for the hydrogel damper at 27°C (pink) and 45°C (red) and for representative damping materials [PDMS (green), Theta 7 (blue), D3O (purple), and 55% cornstarch solution (black)] (n = 5 samples, mean ± SD). (**B**) Relaxation time changes and transition frequency changes with temperature (27° and 45°C) for the hydrogel damper and other damping materials. (**C**) Amplitude difference between peak frequency and 100 Hz for the hydrogel damper at 27° and 45°C and other damping materials. (n = 5 samples, mean ± SD). (**D**) Damping factors for bandpass filtering by the hydrogel damper and other damping materials. (n = 5 samples, mean ± SD). (**E**) Bandwidth comparison of hydrogel damper at 27°C and 45°C and other damping materials. (**F**) Energy damping diagram of the hydrogel damper and other materials (*30*).

than those of the other damping materials (Fig. 3D). Thus, in the case of the hydrogel damper, the damping bandwidths are narrow (80.59 Hz at 27°C and 169.3 Hz at 45°C), whereas other dampers show broad bandwidths (Fig. 3E). Among a broader range of material types in an Ashby plot, the hydrogel damper is located close to conventional hydrogels, near the kilopascal range (Fig. 3F). The hydrogel damper surpasses the tan δ over the elastomers and polymers with the ranges of the kilopascal modulus ranges.

We applied the hydrogel damper into bioelectronics for continuous biosignal detection without signal processing. The amount of

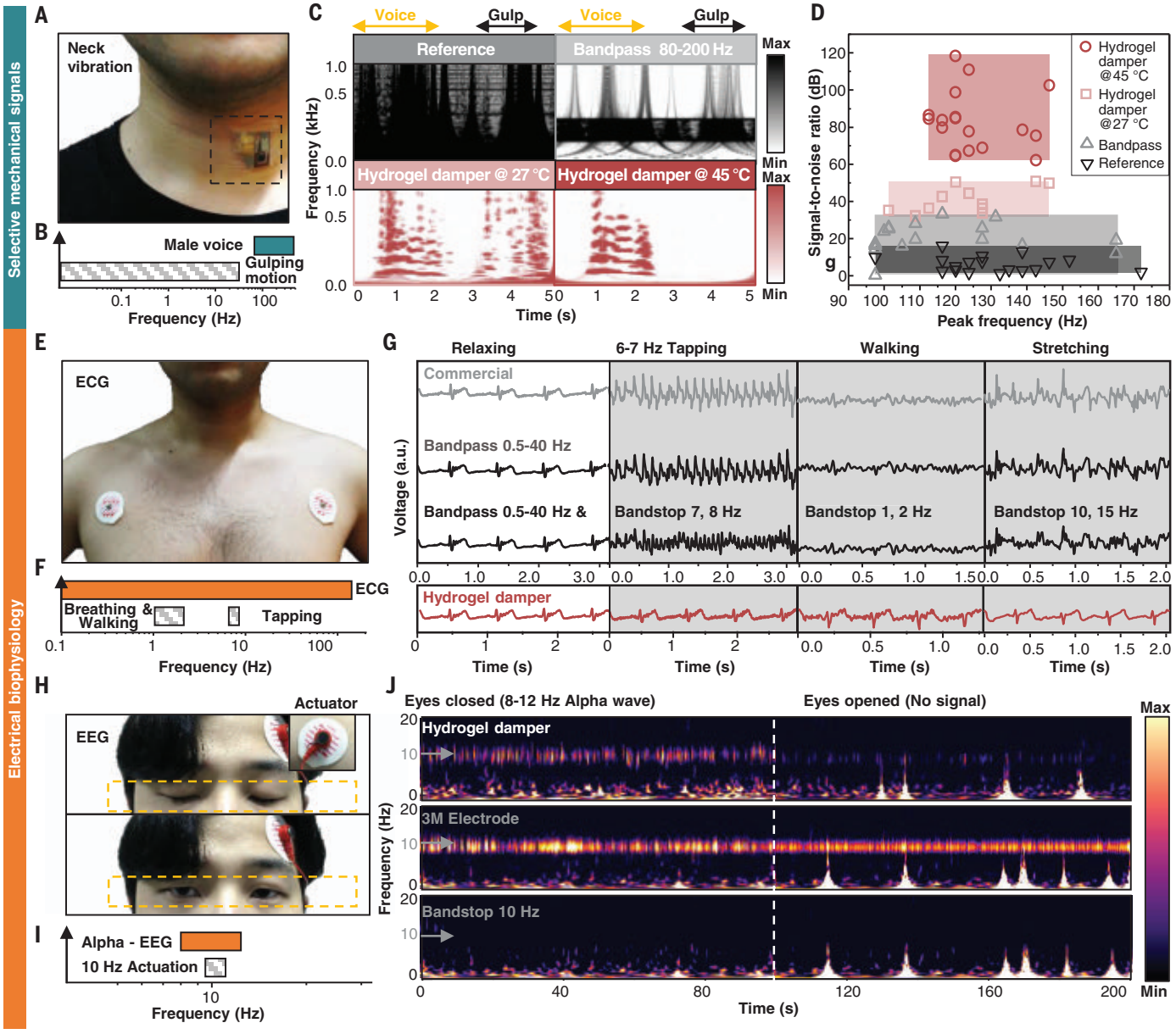


Fig. 4. Demonstration of dynamic noise-damping by using the hydrogel damper for high SNR detection of biophysiological signals. (**A**) Demonstration of human speech recording based on neck vibration by using wireless mechanical sensing with the hydrogel damper and PDMS as a reference. (**B**) Frequency ranges of a male voice and a swallowing motion (~20 Hz). (**C**) Comparisons of auditory spectrograms measured (top) without and (bottom) with the hydrogel damper during the pronunciation of "Viva la vida" and the swallowing motion. Black and gray spectra show reference and bandpass-filtered signals. The acoustic neck vibrations were detected with a reference crack-based sensor on a piece of PDMS and the hydrogel damper at 27°C and 40°C. For comparison, the signals from the reference crack-based sensor were treated with bandpass filtering from 80 to 200 Hz. (**D**) SNR comparison for the detected voice at each peak frequency for the hydrogel damper and reference (PDMS) (n = 18 samples). (**E**) Potential application in ECG detection on the skin with the hydrogel damper and a commercial 3M electrode. (**F**) Frequency ranges of ECG, tapping, breathing, and walking signals. (**G**) ECG signals measured with a commercially available 3M electrode, after bandpass filtering of 0.5 to 40 Hz and (top) additional bandstop filtering and (bottom) using the hydrogel damper under mechanical noise application: tapping, breathing, and walking. (**H**) Another potential application in EEG detection under 10-Hz noise applied with an actuator. (**I**) Typical frequency ranges of EEG and 10-Hz actuation signals. (**J**) Wavelet analysis of EEG signals with (middle) a 3M electrode, (bottom) a commercial electrode with a bandstop centered at 10 Hz, and (top) the hydrogel damper during eye closing and opening. The alpha wave region (8 to ~12 Hz) is targeted under external mechanical noise (10 Hz).

water passivated in the hydrogel damper that is laminated on the different skin spots (chest, neck, wrist, and forehead) does not change significantly (figs. S9 and S11). First, to separate target mechanical signals from low-frequency dynamic noise, we assembled the hydrogel damper with an ultrasensitive crack-based strain sensor (Fig. 1I and fig. S12) (*27*, *28*). One example would be to acquire speech information without dynamic noises deployed on the neck (Fig. 4A). The frequency of neck vibration during speech is usually at least 100 Hz, whereas physical noise arises from motions such as swallowing and skin tension due to movement, which usually has frequencies under 30 Hz (Fig. 4B) (*22*, *23*). A male volunteer spoke "Viva la vida" after attaching the integrated sensor to his neck (Fig. 4C, yellow arrows), and then the volunteer performed a swallowing motion to create dynamic noise (Fig. 4C, black arrows). Even if the sensor has low-frequency dynamic noise induced by the tension of the skin or movement, it contributes over a wide frequency range because the noise can create various scales of friction (*29*). A Morelet wavelet transforms analysis shows that using the control sensors with PDMS and bandpass filtering with 80 to 200 Hz, the vocal signals and swallowing motion could barely be identified (Fig. 4C, fig. S15, and movie S3). By contrast, the spectrogram of the sensor with the hydrogel damper shows a relatively clear separation of acoustic signals from the swallowing motion signal; even at 40°C with the integrated heater, entire noise signals due to motion disappear (fig. S16). The reference, even with the bandpass filtering at 80 to 200 Hz, shows a broad frequency range with the lowest signal-to-noise ratio (SNR), which means that the target signal acquisition is inaccurate, making it hard to discriminate signal from noise (Fig. 4D). With the hydrogel damper, the peak frequency ranges become narrow, and the SNR increases with the temperature (fig. S17). Another example of acquisition, heart rate signals (0.3 to 4 Hz) on the wrist under the lifting move or tapping (<0.4 Hz), supports that the selective mechanical signal target can be exploited regardless of the mechanical noises (fig. S18).

Because the hydrogel damper can be used as a conductive electrode as well as a damping material, we next explored whether it can be used to obtain electrophysiological signals (ECG and EEG) (Fig. 1H and fig. S13). The hydrogel can be expected to transmit the electrophysiology while damping the external mechanical noises. With the lower impedance of the hydrogel (fig. S13), one demonstration involves ECG signal detection (0.05 to 150 Hz range) measurement by attaching each electrode to the chest of a volunteer (Fig. 4E) while the electrode is tapped (7 to 8 Hz) and the participant is breathing and walking (1 to 2 Hz) (Fig. 4F). When the representative signals before and after mechanical noise application were analyzed, the mechanical noise obscured the ECG waveforms, whereas the ECG signal was detected in the hydrogel case regardless of mechanical noise (Fig. 4G). Even bandpass filtering of 0.5 to 40 Hz and additional arbitrary bandstop filtering were ineffective in each case because the noise frequency was within the frequency ranges of the target signals. By contrast, the hydrogel damper results show representative signals that were stable and clear regardless of external mechanical noise (fig. S19). During continuous biophysiology detection over a full day, the mechanical noise-damping capability was not degraded (fig. S20).

We also measured EEG alpha wave signals (8 to 12 Hz) when the eyes are closing by attaching the electrodes to a volunteer's forehead, together with an actuator that generates 10 Hz mechanical noise (Fig. 4, H and I). The wavelet spectrogram shows that both the hydrogel damper and commercial electrodes detect electrical signals at ~9 to 12 Hz when the volunteer closes their eyes (Fig. 4J). When the eyes are open, in which case no signals should appear at 9 to 12 Hz, continuous 10-Hz noise was detected in the commercial electrode cases, and even bandstop filtering of 10 Hz was inefficient because it distorted the signals. However, the hydrogel damper damps the 10-Hz noise selectively, and intermittent signals at 9 to 12 Hz are detected by the hydrogel damper during natural eye closing, indicating that stable electrical signal acquisition is achieved regardless of the noise (fig. S21).

Selective frequency damping with the viscoelastic material minimizes mechanical noise and enables the detection of biophysiological signals with a high SNR under noisy conditions. Rather than signal processing after the interruption by the mechanical noises, selective frequency damping by the material itself would be more effective to acquire clear signals. The strategy for the engineering of relaxation time on demand becomes essential under the various environmental conditions and needs to be further improved in the future work—as well as the temperature controls, which may require additional form factors in whole devices, and the skin can suffer from the temperature controls. In this way, we propose that the real-time application of soft bioelectronics that does not require a signal-processing step can be accelerated with viscoelastic soft materials compared with that of rigid wearable electronics.

REFERENCES AND NOTES

1. D.-H. Kim *et al.*, *Science* **333**, 838–843 (2011).
2. S. Lee *et al.*, *Science* **370**, 966–970 (2020).
3. J. Lee *et al.*, *Nat. Electron.* **4**, 291–301 (2021).
4. T.-i. Kim *et al.*, *Science* **340**, 211–216 (2013).
5. J. Deng *et al.*, *Nat. Mater.* **20**, 229–236 (2021).
6. C. S. Boland *et al.*, *Science* **354**, 1257–1260 (2016).
7. E. Song, J. Li, S. M. Won, W. Bai, J. A. Rogers, *Nat. Mater.* **19**, 590–603 (2020).
8. S. P. Lacour, G. Courtine, J. Guck, *Nat. Rev. Mater.* **1**, 16063 (2016).
9. M. Baumgartner *et al.*, *Nat. Mater.* **19**, 1102–1109 (2020).
10. O. A. Araromi *et al.*, *Nature* **587**, 219–224 (2020).
11. A. Pena-Francesch, H. Jung, M. C. Demirel, M. Sitti, *Nat. Mater.* **19**, 1230–1235 (2020).
12. I. You *et al.*, *Science* **370**, 961–965 (2020).
13. B. C.-K. Tee *et al.*, *Science* **350**, 313–316 (2015).
14. K. Lee *et al.*, *Nat. Biomed. Eng.* **4**, 148–158 (2020).
15. C. J. De Luca, L. D. Gilmore, M. Kuznetsov, S. H. Roy, *J. Biomech.* **43**, 1573–1579 (2010).
16. F. G. Barth, *A Spider's World: Senses and Behavior* (Springer, 2002).
17. S. L. Young *et al.*, *Acta Biomater.* **10**, 4832–4842 (2014).
18. J. F. Vincent, *Compos., Part A Appl. Sci. Manuf.* **33**, 1311–1315 (2002).
19. L. Xu *et al.*, *Sci. Adv.* **5**, eaau3442 (2019).
20. K. Liu *et al.*, *J. Am. Chem. Soc.* **143**, 1162–1170 (2021).
21. E. K. Antonsson, R. W. Mann, *J. Biomech.* **18**, 39–47 (1985).
22. K. L. Venkatachalam, J. E. Herbrandson, S. J. Asirvatham, *Circ. Arrhythm. Electrophysiol.* **4**, 965–973 (2011).
23. S. Brage, N. Brage, P. W. Franks, U. Ekelund, N. J. Wareham, *Eur. J. Clin. Nutr.* **59**, 561–570 (2005).
24. T. McLeish, *Adv. Phys.* **51**, 1379–1527 (2002).
25. M. Tang *et al.*, *ACS Omega* **2**, 2214–2223 (2017).
26. J. A. Plaizier-Vercammen, E. Lecluse, P. Boute, R. E. De Neve, *Dent. Mater.* **5**, 301–305 (1989).
27. D. Kang *et al.*, *Nature* **516**, 222–226 (2014).
28. B. Park *et al.*, *Adv. Mater.* **28**, 8130–8137 (2016).
29. T. Cristiani *et al.*, *Tribol. Lett.* **66**, 149 (2018).
30. A. P. Unwin *et al.*, *Sci. Rep.* **8**, 2454 (2018).

ACKNOWLEDGMENTS

The authors thank S. J. Kwon (SKKU) for helpful discussion on the theoretical modeling; D. S. Hwang (Postech) for the chitosan hydrogel; D. Kim (SKKU) for the high-speed camera; and H.-J. Chung (University of Alberta), H. Lee (KAIST), B. Jin, and S. Jo for helpful discussion on the hydrogel. **Funding:** This work was supported by National Research Foundation of Korea (NRF) grants funded by the Korean government (MSIT) (NRF-2019R1I1A2A01061966, NRF-2020M3C1B8016137, and NRF-2019M3C7A1032076) and by the Technology Innovation Program (20013794, Center for Composite Materials and Concurrent Design) funded by the Ministry of Trade, Industry & Energy (MOTIE, Korea). B.P. was partially supported by Korea Institute for Advancement of Technology (KIAT) grant funded by the Korea Government (MOTIE) [P0017305, Human Resource Development Program for Industrial Innovation (Global)]. **Ethics statement:** All human subject experiments conducted with the volunteers were preapproved by the human subject review board of Sungkyunkwan University (ISKKU 2019-06-015). **Author contributions:** B.P. and T.-i.K. designed the experiments and led this work. B.P. and S.P. synthesized the hydrogel damper. B.P., J.O., and S.P. investigated the mechanical properties of the hydrogel damper. B.P., W.J., C.J., S.C., and Y.J.J. investigated the properties of gelatin and chitosan. B.P. and J.H.S. implemented the detection of biophysiological signals with the hydrogel damper. All authors discussed and shared the results and ideas. B.P. and T.-i.K. wrote the paper. **Competing interests:** T.-i.K, B.P., and S.P. are inventers on a patent application related to this work (KR 10-2020-0061065, 12 May 2021). The authors declare no other competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abj9912
Materials and Methods
Supplementary Text
Figs. S1 to S21
Table S1 and S2
References (*31*–*43*)
Movies S1 to S4

Submitted 16 June 2021; accepted 31 March 2022
10.1126/science.abj9912

EVOLUTION

Idiosyncratic epistasis leads to global fitness–correlated trends

Christopher W. Bakerlee[1,2,3]†, Alex N. Nguyen Ba[1,2,4,5]†, Yekaterina Shulgina[3], Jose I. Rojas Echenique[1,6], Michael M. Desai[1,2,7,8]*

Epistasis can markedly affect evolutionary trajectories. In recent decades, protein-level fitness landscapes have revealed extensive idiosyncratic epistasis among specific mutations. By contrast, other work has found ubiquitous and apparently nonspecific patterns of global diminishing-returns and increasing-costs epistasis among mutations across the genome. Here, we used a hierarchical CRISPR gene drive system to construct all combinations of 10 missense mutations from across the genome in budding yeast and measured their fitness in six environments. We show that the resulting fitness landscapes exhibit global fitness–correlated trends but that these trends emerge from specific idiosyncratic interactions. We thus provide experimental validation of recent theoretical work arguing that fitness-correlated trends can emerge as the generic consequence of idiosyncratic epistasis.

Epistatic interactions have important consequences for the design and evolution of genetic systems (*1*–*3*). Much work in recent decades has studied these interactions by measuring empirical fitness landscapes, most often at "shallow" depth for genome-scale studies (e.g., by quantifying pairwise but not higher-order epistasis between all gene deletions or mutations) or at "narrow" breadth (e.g., complete landscapes at the scale of small select regions in single genes such as by quantifying all orders of epistatic interactions among few amino acid residues) (*4*–*18*). These studies have often found many epistatic interactions among specific mutations at both lower orders (i.e., among few mutations) and higher orders (i.e., among many mutations). These reflect particular biological and physical interactions among the mutations involved; following recent work (*19*, *20*), we refer to them as "idiosyncratic" epistasis because they involve the specific details of these mutations. Overall, this body of work has highlighted the potential for epistasis to create historical contingency that tightly constrains the distribution of adaptive trajectories accessible to natural selection.

By contrast, other work examining adaptive trajectories that implicate loci across the genome has found patterns of apparently "global" epistasis, in which the fitness effect of a mutation varies systematically with the fitness of the genetic background on which it occurs (*21*–*28*). Typically, this manifests as either diminishing returns for beneficial mutations or increasing costs for deleterious mutations, with mutations having a less positive or more negative effect on fitter backgrounds. These consistent patterns of global epistasis may give rise to the dominant evolutionary trend of declining adaptability, and in contrast to the complexity of idiosyncratic interactions, they suggest that historical contingency could be less critical in constraining adaptive trajectories (*29*).

Despite their importance, these dual descriptions of epistasis have not been satisfactorily unified. In one view, global epistasis results from nonspecific fitness-mediated interactions among mutations (*24*). Such interactions may, for example, emerge from the topology of metabolic networks, which generates overall patterns of diminishing returns and increasing costs that eclipse the specific details of idiosyncratic interactions (*30*). By contrast, other recent theoretical work has proposed an alternative view, hypothesizing that apparent fitness-mediated epistasis can instead emerge as the generic consequence of idiosyncratic interactions if they are sufficiently numerous and widespread (*19*, *20*). These two models have substantially different implications for the structure of fitness landscapes, which in turn influence our expectations of the repeatability and predictability of evolution and of the effect of chance and contingency on adaptation at both the genotypic and phenotypic levels. Thus, this dichotomy plays a central role in our understanding of how epistasis affects evolutionary dynamics.

Thus far, however, empirical work has been unable to distinguish between these perspectives. The key difficulty is that testing these ideas requires both depth and breadth: We must analyze landscapes involving enough loci that we sample idiosyncratic interactions that can potentially give rise to overall fitness-mediated trends, and we must survey possible combinations of these mutations at sufficient depth to quantify the role of higher-order interactions (including potential "global" nonspecific fitness-mediated interactions). Larger landscapes are also necessary to reduce the influence of measurement error on the inference of epistasis and analysis of fitness-correlated trends (FCTs) (see the supplementary materials, section 6.3). Achieving this depth and breadth is technically challenging because it requires us to synchronize many mutations across the genome.

Here, we overcame this challenge by developing a method that exploits Cre-Lox recombination to create a combinatorially expanding CRISPR guide-RNA (gRNA) array in *Saccharomyces cerevisiae*, which allowed us to iteratively generate mutations at distant loci through a gene-drive mechanism (Fig. 1A). Briefly, strains of opposite mating type containing inducible Cre recombinase and *SpCas9* genes were mutated at one of two loci (*A* or *B*), and DNA encoding gRNAs specific to the wild-type (WT) alleles at these loci were integrated into their genomes (fig. S1). After mating to produce a diploid heterozygous at *A* and *B*, we induced a gene drive to make the loci homozygous. This began with expressing Cas9 and generating gRNA-directed double-strand breaks at the WT *A* and *B* alleles. These breaks were then repaired by the mutated regions of homologous chromosomes, making the diploid homozygous at these loci with at least 95% efficiency. Simultaneously, we expressed Cre to induce recombination that brought gRNAs into physical proximity on the same chromosome by way of flanking Lox sites using a strategy similar to that described previously (*31*) (Fig. 1B). We then sporulated diploids and selected haploids bearing the linked gRNAs from both parents. In parallel, we performed this process with "pseudo-WT" versions of these loci, which contain synonymous changes that abolish gRNA recognition but lack the nonsynonymous change of interest. This created a set of four strains, with all possible genotypes at loci *A* and *B*. Concurrently, we created separate sets of four strains with all possible genotypes at other pairs of loci (e.g., *C* and *D*).

By iterating this process, we could rapidly assemble an exponentially expanding, combinatorially complete genotype library. We mated separate sets of four genotypes bearing all combinations of mutations at two loci each in an all-against-all cross, drove their mutations, recombined their gRNAs, and sporulated to produce a 16-strain library bearing all four-locus mutation combinations. Repeating these steps in a third cycle with two four-locus libraries of opposite mating type yielded a 256-strain, eight-locus library, and a complete landscape of up to 16 mutations (2^{16} strains) could be constructed in just four cycles.

[1]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA. [2]Quantitative Biology Initiative, Harvard University, Cambridge, MA, USA. [3]Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA, USA. [4]Department of Cell and Systems Biology, University of Toronto, Toronto, Ontario, Canada. [5]Department of Biology, University of Toronto Mississauga, Mississauga, Ontario, Canada. [6]Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada. [7]NSF-Simons Center for Mathematical and Statistical Analysis of Biology, Harvard University, Cambridge, MA, USA. [8]Department of Physics, Harvard University, Cambridge, MA, USA.
*Corresponding author. Email: mdesai@oeb.harvard.edu
†These authors contributed equally to this work

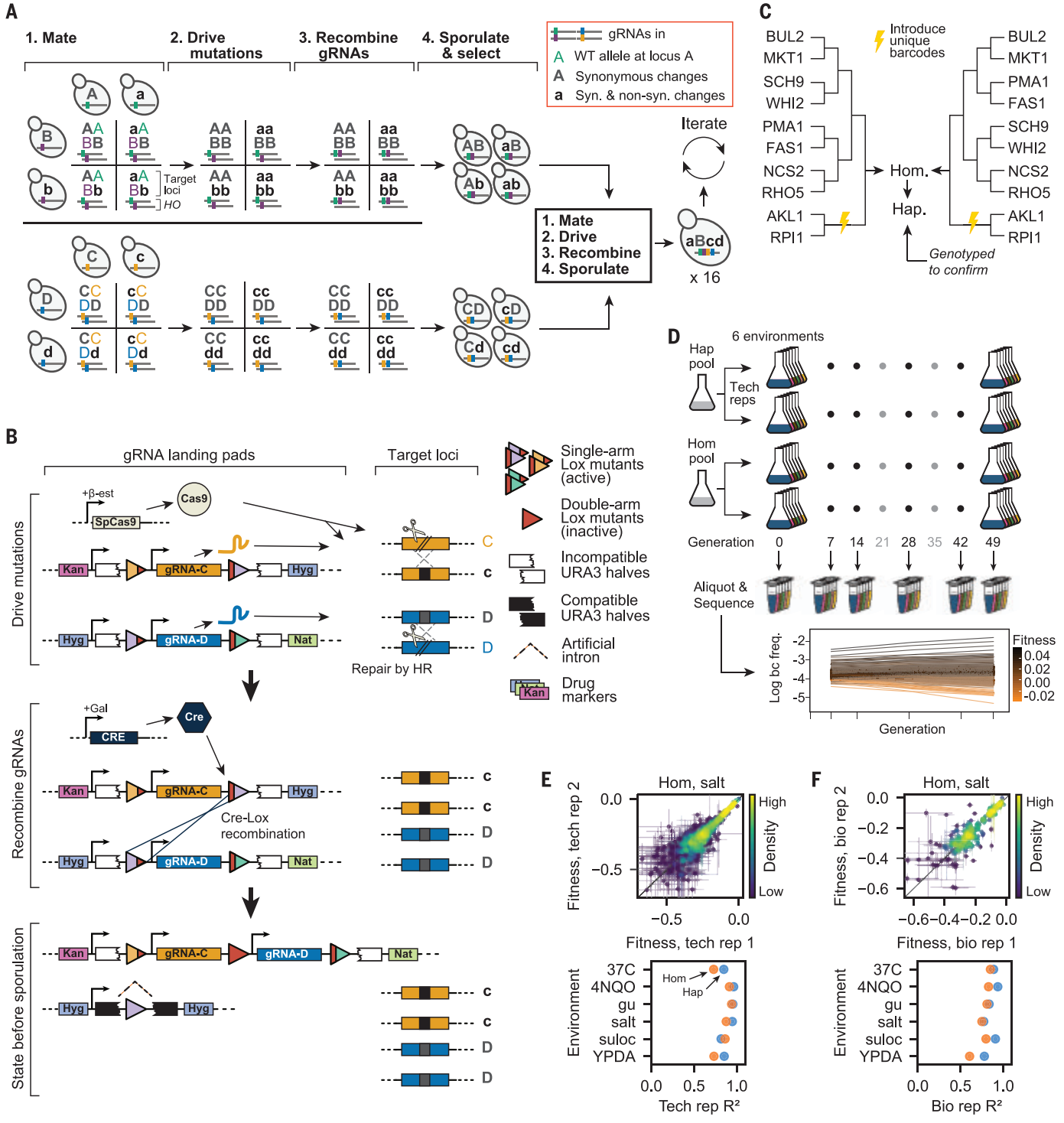


Fig. 1. Recombining CRISPR-gene drive system. (**A**) Experimental design. Strains of opposite mating type carrying known mutations and corresponding gRNAs mate to form heterozygous diploids. Cas9 expression "drives" these mutations, and site-specific recombination links gRNAs. Homozygous diploids are sporulated, haploids with linked gRNAs are selected, and the process repeats, incorporating exponentially increasing numbers of mutations. (**B**) Recombining gene drive system. gRNAs targeting heterozygotic loci are flanked by selection markers and two of three orthogonal Lox sites (colored triangles), which are inactivated through recombination (red triangles). Cas9 "drives" targeted mutations, whereas Cre-Lox recombination brings like markers to the same chromosome and activates a *URA3* gene interrupted by an artificial intron. After sporulation, the chromosome with gRNAs is selected using the markers of interest and the other is counterselected using 5-fluoroorotic acid (5-FOA). (**C**) Cross-design. A complete fitness landscape is produced in parallel by distinct cross designs that yield final homozygous diploids and haploids in biological replicates with unique DNA barcodes. (**D**) Bulk-fitness assays. Pooled strains are assayed in replicate for competitive fitness in several environments by sequencing barcodes to obtain strain frequencies over time. (**E**) Repeatability of technical replicate competitive fitness measurements. (**F**) Repeatability of biological replicate competitive fitness measurements.

We sought to use this method to construct a complete fitness landscape that would shed light on the structure of epistasis: Are FCTs primarily the product of a global coupling of mutations through fitness or do they emerge as the consequence of idiosyncratic epistasis? We surveyed studies of natural variation [e.g., (*32*–*36*)] and experimental evolution [e.g., (*37*–*39*)] to identify mutations potentially relevant to adaptation in the laboratory strain.

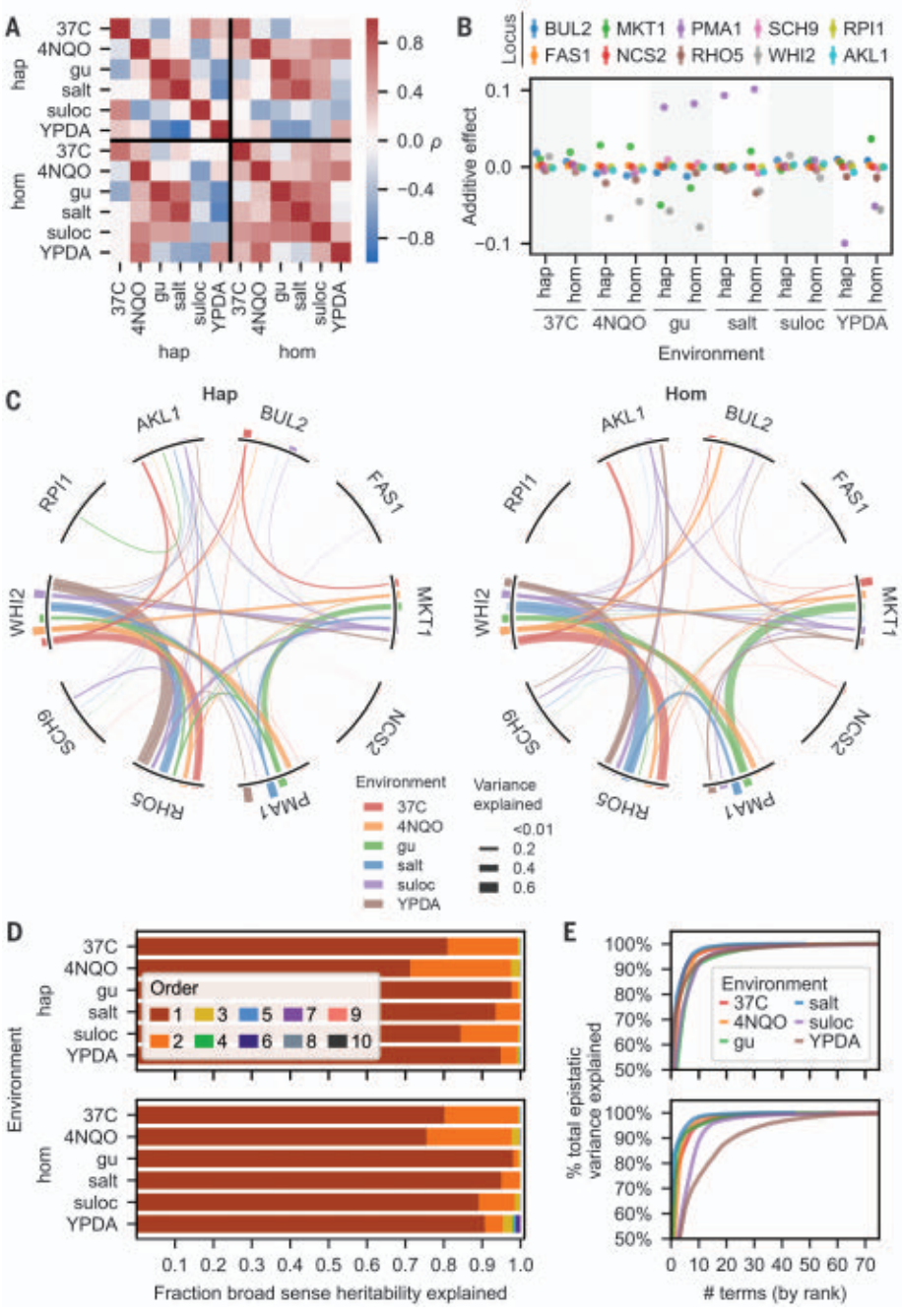


Fig. 2. Fitness landscapes. (**A**) Correlation in observed fitness (upper right) and predicted fitness (from inferred model, lower left; see the supplementary materials, section 5.1) across ploidies and environments. (**B**) Background-averaged additive effect of each locus across ploidies and environments. Error bars represent 95% confidence intervals. (**C**) Background-averaged pairwise epistatic effects between loci across ploidies and environments. Weights of edges connecting loci represent the proportion of pairwise variance explained by each interaction. Heights of bars on the perimeter correspond to the proportion of additive variance explained by each locus in each environment. (**D**) Variance partitioning of broad-sense heritability from additive and epistatic orders across ploidies and environments. (**E**) Cumulative distribution of the epistatic variance explained by rank-ordered epistatic terms of all orders.

We selected a set of mutations that sample a wide range of cellular functions, such as membrane stress response, mitochondrial stability, and nutrient sensing. Our goal in making this choice was to maximize fitness variance while minimizing pathway-specific idiosyncratic interactions. Alternative choices of mutations, particularly if they were focused on a specific protein or pathway (or limited to those that accumulated along the line of descent in a single lineage), might exhibit very different patterns of epistasis, which would be characteristic of the particular details of that specific protein or pathway (or that specific adaptive trajectory). However, our goal here was to analyze potentially global patterns of epistasis among mutations across the genome that are relevant to fitness in a variety of conditions and thus represent an overall fitness landscape for the laboratory strain.

We thus implemented our gene-drive system to construct a near-complete landscape spanning 10 missense mutations in 10 genes (including essential genes) on eight chromosomes: *AKL1* (S176P), *BUL2* (L883F), *FAS1* (G588A), *MKT1* (D30G), *NCS2* (H71L), *PMA1* (S234C), *RHO5* (G10S), *RPI1* (E102D), *SCH9* (P220S), and *WHI2* (L262S) (Fig. 1C and table S1). We found that a landscape of about this size is required to distinguish the two models (see the supplementary materials, section 6.3). Immediately before the final mating cycle, all strains were transformed with a unique DNA barcode next to the *LYS2* locus to enable high-throughput, sequencing-based competitive fitness assays (figs. S2 and S3). All strains in each replicate haploid library were genotyped at all 10 loci to confirm the presence of the desired alleles (this step also ensures presence in the diploid libraries). We excluded strains in which gene drive failure led to improper genotypes, such that ultimately 875 out of 1024 (85.4%) genotypes remained (468 were present in just one library, and 407 were present in both libraries). We also performed whole-genome sequencing of 96 randomly selected strains to rule out pervasive aneuploidies or influential but spurious background mutations. One aneuploidy was identified, and three spurious background mutations were observed at >5% frequency. Subsequent analysis showed that these were unlikely to systematically influence our findings (table S2 and supplementary materials, section 5.1).

To obtain fitness landscapes, we conducted replicate bulk barcode–based fitness assays on both pooled haploid and homozygous diploid versions of the genotype library in six distinctly stressful media environments: yeast extract peptone dextrose (YPD) + 0.4% acetic acid, YPD + 6 mM guanidium chloride, YPD + 35 μM suloctidil, YPD at 37°C, YPD + 0.8 M NaCl, and synthetic dextrose (SD) + 10 ng/ml 4-nitroquinoline 1-oxide (4-NQO) (Fig. 1D). For each of 7 days, pools were allowed seven

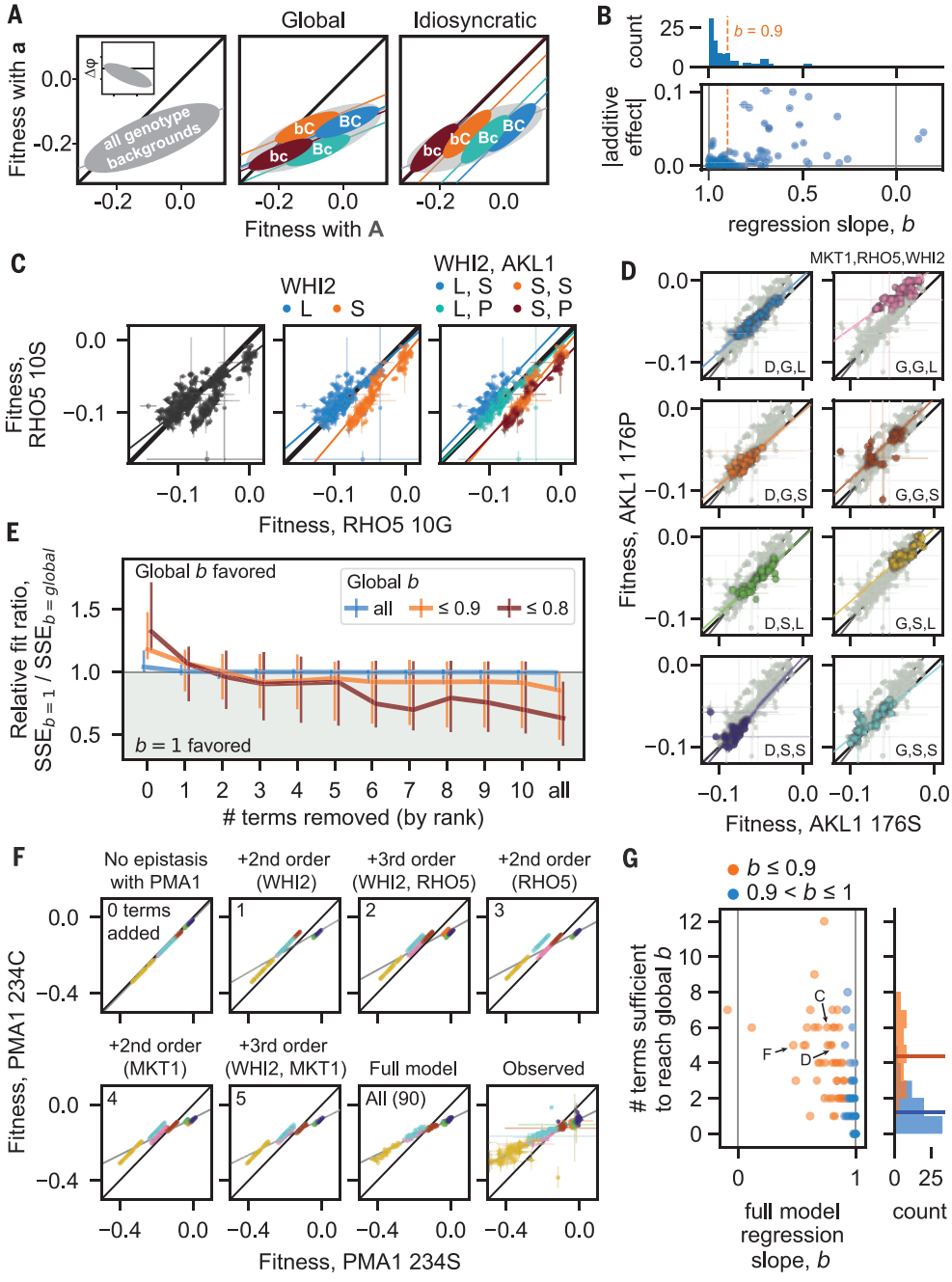


Fig. 3. FCTs. (**A**) Schematic contrasting how global or idiosyncratic epistasis could produce FCTs. Inset shows FCT analyzed as the effect of a mutation ($\Delta\varphi$) on backgrounds of different fitnesses. (**B**) Histogram and scatterplot of regression slopes (b) between φ_{Mut} and φ_{WT}, and corresponding absolute additive effects of mutations. Polarity was adopted such that $b \leq 1$. Total error bar length is twice the standard error of the slope. (**C**) Fitness effect of *RHO5* mutation (G10S) (φ_{Mut} versus φ_{WT}) in all haploid backgrounds at 37°C (left) and partitioned by genotypes at *WHI2* (L262S) (middle) and *WHI2* and *AKL1* (S176P) (right). Initial $SSE_{b=1}/SSE_{b=global}$ is 1.21. (**D**) Fitness effect of *AKL1* mutation in all homozygote backgrounds in the suloctidil environment, partitioned by genotypes at *MKT1* (D30G), *RHO5*, and *WHI2*. Initial $SSE_{b=1}/SSE_{b=global}$ = 1.31. (**E**) Median relative fit ratio between regressions with fixed slope of b = 1 and b = global as function of number of epistatic terms removed from observed phenotypes. Vertical lines represent interquartile range. Polarity was adopted such that $b \leq 1$. (**F**) Inferred fitness effect of *PMA1* S234C mutation in 4-NQO environment across all haploid backgrounds. Epistatic terms interacting with *PMA1* are completely removed from genotype fitnesses, then added back sequentially (from largest to smallest). Bottom-right: full-model (inferred) and observed genotype fitnesses, respectively. Gray line is regression slope. (**G**) Scatterplot and histograms of FCT regression slopes for all data and the number of epistatic terms sufficient to recapitulate them. Horizontal lines in the histogram indicate means. Arrows and letters indicate populations presented in previous panels. Polarity was adopted such that $b \leq 1$.

generations of growth, and aliquots were sampled and sequenced at the barcode locus at generations 7, 14, 28, 42, and 49. We estimated the relative fitness of each genotype from changes in barcode frequencies through time, achieving consistent measurements across technical and biological replicates (Fig. 1, E and F, and fig. S4). From these data, we inferred the background-averaged additive and epistatic effects of each mutation and combination of mutations, respectively (using LASSO regularization; see the supplementary materials).

We found that our six environments yielded substantially different landscapes, as demonstrated by the relatively low between-environment correlations of genotype fitnesses (Fig. 2A), the additive effects of each mutation (Fig. 2B), and the pairwise interactions between them (Fig. 2C). Haploid and homozygous diploid landscapes were largely correlated, but there were several notable exceptions, particularly in the suloctidil environment (Fig. 2, A and B). Although some pairwise interactions remain roughly constant in strength even as the corresponding additive effects varied considerably (e.g., *RHO5* and *WHI2*), most waxed and waned across environments (Fig. 2C). Nevertheless, the overall contribution from different epistatic orders showed some similarities across ploidies and environments (the magnitudes did differ; Fig. 2D), with additive and pairwise terms explaining most of the variance in the data, third-order terms contributing minorly, and the remaining orders making little difference, consistent with earlier studies (*40*). Across all epistatic orders, inferred effects were highly skewed, with a small number of terms explaining disproportionate variance (Fig. 2E).

We next sought to investigate potential patterns of global fitness–mediated epistasis. To do so, for each locus in each ploidy and environment, we plotted the fitness of a genotype

with the mutated allele (φ_{Mut}) against the fitness of the same genotype with the WT allele (φ_{WT}). A regression slope (b) different from 1 in these plots signifies an FCT (Fig. 3A, left; see the supplementary materials). Some previous work instead plotted the fitness effect of a mutation ($\Delta\varphi$) as a function of background fitness (φ_{WT}). The advantage of our formulation here is that it does not privilege a specific allele as the "WT." Instead, regression in our plots translates intuitively when reversing direction to treat the reversion as the mutation: $b_{rev} = 1/b_{orig}$ by weighted-total least squares (see the expanded discussion in the supplementary materials; figs. S5 to S8).

We found that FCTs were common in our landscapes: Across all ploidies, environments, and loci, ~44% of regression slopes deviated substantially from 1 (i.e., $b \leq 0.9$ or $b \geq 0.9^{-1}$; these deviations were all significant; Fig. 3B, histogram, and figs. S13 and S14). However, FCTs were not universal for fitness-affecting mutations: Of the 49 examples across ploidies and environments of mutations with additive effects of magnitude ≥ 0.5%, 18 were associated with $0.9 < b < 0.9^{-1}$ (Fig. 3B).

By partitioning background genotypes by the presence or absence of specific mutations, we could determine whether FCTs were truly "global" (i.e., whether they transcended these partitions and any corresponding idiosyncratic interactions; Fig. 3A, middle) or were instead fundamentally idiosyncratic (i.e., they emerged from regression across partitions shifted in φ_{Mut} versus φ_{WT} space by sparse interactions with specific background loci; Fig. 3A, right). When we partitioned FCTs by the presence or absence of interacting mutations in the background, we found several instances in which the idiosyncratic model clearly explained the FCT. For example, the effect of the G10S mutation in *RHO5* at 37°C exhibited a clear FCT (b = 0.76) (Fig. 3C). However, we could partition points by the presence of interacting *WHI2* and *AKL1* alleles in the background. Doing so showed that pairwise interactions with these alleles caused systematic shifts in φ_{10S} versus φ_{10G} space, with each partition assuming a slope near 1. Thus, over a range of background fitnesses, an FCT in the effect of the G10S emerged from these specific idiosyncratic interactions (Fig. 3C and fig. S11). In the case of the homozygous *AKL1* S176P mutation in suloctidil, we observed a similar decomposition of an FCT (b = 1.29) when partitioning genotypes according to the presence of three interacting loci in the background (*MKT1*, *RHO5*, and *WHI2*) (Fig. 3D and fig. S11). However, in other cases, it was less clear whether the FCT could be partitioned in this way, and because deeper partitions tend to reduce background fitness variance and limit our confidence in regression slopes, a different approach was required to characterize the extent to which idiosyncratic terms caused FCTs across our data.

To investigate this, we analyzed the effect of removing specific idiosyncratic epistatic terms on the overall FCTs. To do so, for each focal locus (in each ploidy and environment), we first calculated the weighted sum of squared errors (SSE) (*41*) of observed fitnesses from the global regression line ($SSE_{b=global}$) and from a fitted line of slope 1 ($SSE_{b=1}$, which corresponds to no FCT). We then set the largest epistatic term to zero and recalculated the expected fitness of each resulting genotype (assuming that all other terms and residuals were nonzero), again obtaining both $SSE_{b=global}$ and $SSE_{b=1}$. If the FCT arose from a global effect, then we would expect that $SSE_{b=global}$ would be less than $SSE_{b=1}$ even as terms were removed. Instead, we found that after removing the effect of just a few terms, a regression with a fixed slope of b = 1 typically fit the data better than the b = global FCT slope (with the FCT threshold set to $b \leq 0.9$ or 0.8; Fig. 3E and fig. S11), approaching the fit of an unconstrained regression that minimizes SSE (i.e., the final slope approaches 1; fig. S10). This indicates that the apparent FCT arises from these few idiosyncratic interactions, even for global slopes very different from 1. Although we also documented cases in which b = global fit the data better than b = 1 even after removing many terms, we expect that most, if not all, of these instances may have been due to measurement error because they tended to arise in ploidies and environments in which the data were noisier (fig. S17).

To further evaluate whether idiosyncratic interactions between these mutations were sufficient to generate FCTs, we performed the converse analysis, this time with genotype fitnesses as predicted by our model of additive and idiosyncratic epistatic terms. Instead of removing the effects of epistatic terms one at a time, we first stripped from the model all interactions involving the focal locus, yielding perfectly linear points of slope 1 when plotting φ_{Mut} versus φ_{WT}. We then added interactions one-by-one to our fitness prediction, from largest to smallest, and examined the resulting slopes. As shown in Fig. 3F for the haploid *PMA1* S234C mutation in 4-NQO, adding just a few terms associated with three background loci recapitulated a strong FCT. Repeating this analysis with all our mutations showed that, on average, just four idiosyncratic interactions (primarily pairwise) were sufficient to recapitulate the full-model FCTs (a slope within 0.01 of the global slope; Fig. 3G, orange; see the supplementary materials), which is far lower than the total number of inferred terms (median of 53) but represents on average 89% of the potential variance explained that could have been added (fig. S12). Thus, although FCTs are real and likely have important biological consequences, our data demonstrate that apparent fitness-mediated epistasis can readily emerge from very few low-order idiosyncratic interactions.

Because the landscapes that we studied here have no natural polarization (i.e., neither allele is the assumed WT), we cannot comment directly on why earlier studies of global epistasis have more commonly found negative than positive FCTs (when plotting $\Delta\varphi$ versus φ_{WT}). However, this distribution of FCT directions is important because it may underly the ubiquitous trend of declining adaptability observed across laboratory evolution experiments (*29*). The observed bias toward negative trends may arise from asymmetries in the average sign of epistatic interactions between mutations away from extant high-fitness genotypes relative to their reversions, which theory has predicted should arise from idiosyncratic interactions (*19*, *20*). In addition, choosing polarizations at random will lead to more negative than positive FCTs across the full parameter space (see the extended discussion in the supplementary materials).

Regardless of the cause of any asymmetry in the direction of FCTs, our results support recent theoretical arguments that fitness-mediated epistasis can emerge as the generic consequence of widespread idiosyncratic interactions, rather than reflecting a global fitness–mediated coupling of mutations. Indeed, at least in our system, FCTs could arise even from a relatively small number of low-order interactions. We note that landscapes involving other types of variation [e.g., within a single protein or pathway or along the line of descent in a single lineage (*21*)] may exhibit different patterns, although we may expect these scenarios to involve an even stronger role for idiosyncratic interactions. More generally, we emphasize that idiosyncratic epistasis and global fitness–mediated effects are not mutually exclusive, and although FCTs can be explained by the former in our system, in other cases, both effects may contribute. However, our results suggest that nonspecific global epistasis may not be the primary driver of patterns of declining adaptability in laboratory evolution experiments, and this has general implications for the ways in which epistasis constrains evolutionary trajectories.

REFERENCES AND NOTES

1. D. J. Kvitek, G. Sherlock, *PLOS Genet.* **7**, e1002056 (2011).
2. P. A. Romero, F. H. Arnold, *Nat. Rev. Mol. Cell Biol.* **10**, 866–876 (2009).
3. K. M. Flynn, T. F. Cooper, F. B.-G. Moore, V. S. Cooper, *PLOS Genet.* **9**, e1003426 (2013).
4. M. Costanzo *et al.*, *Science* **353**, aaf1420 (2016).
5. A. D. Norris, X. Gracida, J. A. Calarco, *eLife* **6**, e28129 (2017).
6. M. A. Horlbeck *et al.*, *Cell* **174**, 953–967.e22 (2018).
7. K. S. Sarkisyan *et al.*, *Nature* **533**, 397–401 (2016).
8. C. A. Olson, N. C. Wu, R. Sun, *Curr. Biol.* **24**, 2643–2651 (2014).

9. S. Bershtein, M. Segal, R. Bekerman, N. Tokuriki, D. S. Tawfik, *Nature* **444**, 929–932 (2006).
10. N. C. Wu *et al*., *Nat. Commun.* **11**, 1233 (2020).
11. V. O. Pokusaeva *et al*., *PLOS Genet.* **15**, e1008079 (2019).
12. J. I. Jiménez, R. Xulvi-Brunet, G. W. Campbell, R. Turk-MacLeod, I. A. Chen, *Proc. Natl. Acad. Sci. U.S.A.* **110**, 14984–14989 (2013).
13. B. Madan *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2011653118 (2021).
14. N. C. Wu, L. Dai, C. A. Olson, J. O. Lloyd-Smith, R. Sun, *eLife* **5**, e16965 (2016).
15. C. Bank, S. Matuszewski, R. T. Hietpas, J. D. Jensen, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 14085–14090 (2016).
16. J. Domingo, G. Diss, B. Lehner, *Nature* **558**, 117–121 (2018).
17. J. Teyra, A. Ernst, A. Singer, F. Sicheri, S. S. Sidhu, *Protein Sci.* **29**, 433–442 (2020).
18. T. V. Lite *et al*., *eLife* **9**, e60924 (2020).
19. D. M. Lyons, Z. Zou, H. Xu, J. Zhang, *Nat. Ecol. Evol.* **4**, 1685–1693 (2020).
20. G. Reddy, M. M. Desai, *eLife* **10**, e64740 (2021).
21. A. I. Khan, D. M. Dinh, D. Schneider, R. E. Lenski, T. F. Cooper, *Science* **332**, 1193–1196 (2011).
22. H.-H. Chou, H.-C. Chiu, N. F. Delaney, D. Segrè, C. J. Marx, *Science* **332**, 1190–1192 (2011).
23. L. Perfeito, A. Sousa, T. Bataillon, I. Gordo, *Evolution* **68**, 150–162 (2014).
24. S. Kryazhimskiy, D. P. Rice, E. R. Jerison, M. M. Desai, *Science* **344**, 1519–1522 (2014).
25. M. S. Johnson, A. Martsul, S. Kryazhimskiy, M. M. Desai, *Science* **366**, 490–493 (2019).
26. X. Wei, J. Zhang, *Mol. Biol. Evol.* **36**, 1008–1021 (2019).
27. S. Schoustra, S. Hwang, J. Krug, J. A. G. M. de Visser, *Proc. Biol. Sci.* **283**, 20161376 (2016).
28. R. C. MacLean, G. G. Perron, A. Gardner, *Genetics* **186**, 1345–1354 (2010).
29. A. Couce, O. A. Tenaillon, *Front. Genet.* **6**, 99 (2015).
30. S. Kryazhimskiy, *eLife* **10**, e60200 (2021).
31. A. N. Nguyen Ba *et al*., *Nature* **575**, 494–499 (2019).
32. H. Sinha *et al*., *Genetics* **180**, 1661–1670 (2008).
33. S. W. Doniger *et al*., *PLOS Genet.* **4**, e1000183 (2008).
34. E. X. Kwan, E. Foss, L. Kruglyak, A. Bedalov, *PLOS Genet.* **7**, e1002250 (2011).
35. J. S. Bloom, I. M. Ehrenreich, W. T. Loo, T.-L. V. Lite, L. Kruglyak, *Nature* **494**, 234–237 (2013).
36. D. M. Wloch-Salamon, K. Tomala, D. Aggeli, B. Dunn, *G3* **7**, 1899–1911 (2017).
37. B. Szamecz *et al*., *PLOS Biol.* **12**, e1001935 (2014).
38. L. M. Kohn, J. B. Anderson, *Eukaryot. Cell* **13**, 1200–1206 (2014).
39. S. Venkataram *et al*., *Cell* **166**, 1585–1596.e22 (2016).
40. D. M. Weinreich, Y. Lan, J. Jaffe, R. B. Heckendorn, *J. Stat. Phys.* **172**, 208–225 (2018).
41. D. York, N. M. Evensen, M. L. Martínez, J. De Basabe Delgado, *Am. J. Phys.* **72**, 367–375 (2004).
42. C. W. Bakerlee, A. N. Nguyen Ba, Fitness-correlated trends analysis pipeline publication archive (Zenodo, 2022; https://zenodo.org/record/6352707).

ACKNOWLEDGMENTS

We thank the Bauer Core facility at Harvard; G. Reddy, B. Shraiman, and members of the Desai laboratory for experimental assistance and comments on the manuscript; and A. Rego-Costa for help in creating Fig. 2C. Computational work was performed on the Odyssey cluster supported by the Research Computing Group at Harvard University. **Funding:** This work was supported by the Department of Defense (DoD), National Defense Science & Engineering Graduate (NDSEG) Fellowship Program (C.W.B.); the Natural Sciences and Engineering Research Council of Canada (postdoctoral fellowship to A.N.N.B.); Discovery Grant RGPIN-2021-02716 and Discovery Launch supplement DGECR-2021-00117 (A.N.N.B.); the National Human Genome Research Institute of the National Institutes of Health (award F31-HG010984 to Y.S.); the National Science Foundation (grant PHY-1914916 to M.M.D.); and the National Institutes of Health (grant GM104239 to M.M.D.). **Author contributions:** Conceptualization: A.N.N.B., C.W.B., Y.S., J.I.R., M.M.D.; Formal analysis: C.W.B., A.N.N.B.; Funding acquisition: M.M.D.; Investigation: C.W.B., A.N.N.B.; Methodology: A.N.N.B., C.W.B., Y.S., J.I.R.; Resources: M.M.D.; Supervision: M.M.D.; Visualization: C.W.B., A.N.N.B.; Writing – original draft: C.W.B., A.N.N.B., M.M.D.; Writing – review and editing: C.W.B., A.N.N.B., M.M.D. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Raw sequencing data are available at the National Center for Biotechnology Information (NCBI) Sequence Read Archive (accession no. PRJNA815849), and analysis code is available from Github (*42*). All other data are presented in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm4774
Methods and Supplementary Analysis
Figs. S1 to S17
Table S1
References (*43*–*59*)
Data S1 to S3
MDAR Reproducibility Checklist

Submitted 22 September 2021; resubmitted 9 January 2022
Accepted 1 April 2022
10.1126/science.abm4774

CONSERVATION

The critically endangered vaquita is not doomed to extinction by inbreeding depression

Jacqueline A. Robinson[1]*†, Christopher C. Kyriazis[2]*†, Sergio F. Nigenda-Morales[3], Annabel C. Beichman[4], Lorenzo Rojas-Bracho[5,6]*, Kelly M. Robertson[7], Michael C. Fontaine[8,9,10], Robert K. Wayne[2], Kirk E. Lohmueller[2,11]*, Barbara L. Taylor[7]*, Phillip A. Morin[7]*

In cases of severe wildlife population decline, a key question is whether recovery efforts will be impeded by genetic factors, such as inbreeding depression. Decades of excess mortality from gillnet fishing have driven Mexico's vaquita porpoise (*Phocoena sinus*) to ~10 remaining individuals. We analyzed whole-genome sequences from 20 vaquitas and integrated genomic and demographic information into stochastic, individual-based simulations to quantify the species' recovery potential. Our analysis suggests that the vaquita's historical rarity has resulted in a low burden of segregating deleterious variation, reducing the risk of inbreeding depression. Similarly, genome-informed simulations suggest that the vaquita can recover if bycatch mortality is immediately halted. This study provides hope for vaquitas and other naturally rare endangered species and highlights the utility of genomics in predicting extinction risk.

A central question for populations that have undergone severe declines is whether recovery is possible or whether it may be hindered by deleterious genetic factors (*1*). Perhaps the most immediate genetic threat in populations of very small size (<25 individuals) is the deterioration of fitness as a result of inbreeding depression (*2*, *3*). Thus, predicting the threat of inbreeding depression under various genetic and demographic conditions is essential for the conservation of endangered species.

[1]Institute for Human Genetics, University of California, San Francisco, San Francisco, CA, USA. [2]Department of Ecology and Evolutionary Biology, University of California, Los Angeles, Los Angeles, CA, USA. [3]Advanced Genomics Unit, National Laboratory of Genomics for Biodiversity (Langebio), Center for Research and Advanced Studies (Cinvestav), Irapuato, Guanajuato, Mexico. [4]Department of Genome Sciences, University of Washington, Seattle, WA, USA. [5]Comisión Nacional de Áreas Naturales Protegidas/SEMARNAT, Ensenada, Mexico. [6]PNUD-Sinergia en la Comisión Nacional de Áreas Naturales Protegidas, Ensenada, Mexico. [7]Southwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration (NOAA), La Jolla, CA, USA. [8]MIVEGEC, Université de Montpellier, CNRS, IRD, Montpellier, France. [9]Centre de Recherche en Écologie et Évolution de la Santé (CREES), Montpellier, France. [10]Groningen Institute for Evolutionary Life Sciences (GELIFES), University of Groningen, Groningen, Netherlands. [11]Department of Human Genetics, David Geffen School of Medicine, University of California, Los Angeles, Los Angeles, CA, USA.
***Corresponding author. Email: jacqueline.robinson@ucsf.edu (J.A.R.); ckyriazis@g.ucla.edu (C.C.K.); lrojasbracho@gmail.com (L.R.-B.); barbara.taylor@noaa.gov (B.L.T.); klohmueller@g.ucla.edu (K.E.L.); phillip.morin@noaa.gov (P.A.M.)**
†These authors contributed equally to this work.

The critically endangered vaquita porpoise (*Phocoena sinus*), found only in the northernmost Gulf of California, Mexico, has declined from ~600 individuals in 1997 to ~10 individuals at present (*4*). This precipitous decline has been driven by incidental mortality in fishing gillnets (bycatch) (*4*, *5*) (Fig. 1A). Efforts to reduce the intensity of illegal gillnet fishing and implement stronger protections for vaquitas have not been successful, and vaquitas are now considered the most endangered marine mammal (*4*). A recent viability analysis found that the vaquita population could theoretically rebound if bycatch mortality is eliminated (*6*). However, the degree to which genetic factors may prevent a robust recovery is unknown, which has led some to argue that the species is doomed to extinction from genetic threats (*1*, *7*, *8*).

Population viability analysis (PVA) has long been an important tool for modeling extinction risk (*9*). However, it is often challenging to parameterize PVA models for highly endangered species, where information on the potential impact of inbreeding depression is limited. Genomic data offer a potential solution because they can be used to estimate the fundamental genetic and demographic parameters that underlie inbreeding depression. Although the potential applications of genomics in conservation have been widely discussed (*10*, *11*), genomics remain underutilized in

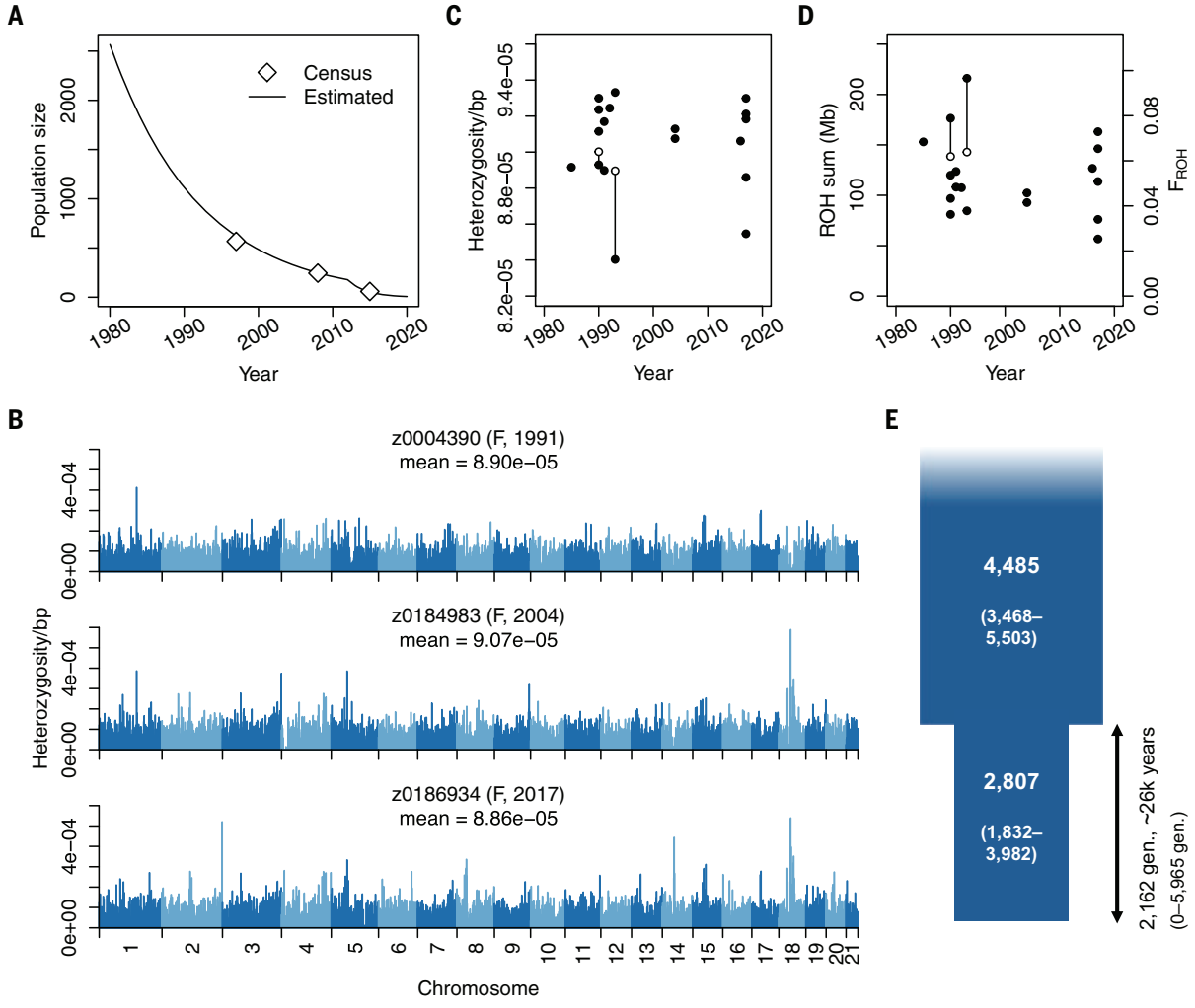


Fig. 1. Vaquita genome-wide diversity and demographic history. (**A**) Model of vaquita census population size based on previous surveys (*5*) shows a marked recent decline. (**B**) Bar plots of per-site heterozygosity in 1-Mb genomic windows in three individuals (one from each sampling period; see fig. S1 for all) show little variability within or between individuals. bp, base pair. (**C** and **D**) Genome-wide heterozygosity (C) and ROH burden (D) are consistent between sampling periods. Lines connect mother-fetus pairs, and open symbols indicate offspring. (**E**) Two-epoch demographic model inferred with ∂a∂i. Parameter 95% confidence intervals are indicated in parentheses. k, thousand; gen., generations.

forecasts of population viability and extinction risk.

To investigate the effect of the vaquita's recent decline and to quantify the species' recovery potential, we sequenced genomic DNA of 19 archival tissue samples to high depth [total n = 20 samples, including the genome from (*12*); mean coverage = 60×] (table S1). Samples were obtained across three time periods: 1985 to 1993, 2004, and 2016 to 2017, spanning approximately three vaquita generations [assuming a generation time of 11.9 years; (*13*)] and an estimated ~99% decline in population size (Fig. 1A) (*5*). All 20 vaquita genomes contain uniformly low heterozygosity [mean = 9.04×10^{-5}, standard deviation (SD) = 2.44×10^{-6} heterozygotes per site; Fig. 1B and fig. S1], consistent with a previous estimate from a single individual (*12*). Additionally, genome-wide diversity appears stable over the sampling period (Fig. 1, B and C), as expected given the short duration of the decline.

We also investigated whether vaquita genomes show signs of recent inbreeding. We found that the mean cumulative fraction of vaquita genomes in long (≥1 Mb) runs of homozygosity (ROH) is 5.42% (SD = 1.7%), which implies a low average inbreeding coefficient of F_{ROH} = 0.05 (Fig. 1D and fig. S2). Furthermore, ROH in our sample are relatively short (mean length, 1.59 to 3.18 Mb), which suggests that they trace to a common ancestor from ~15 to 31 generations ago (178 to 369 years) (*5*). This result indicates that these ROH are a consequence of the vaquita's historically limited population size rather than a consequence of recent inbreeding. Finally, we found limited evidence for close relatives in our dataset, aside from two known mother-fetus pairs (fig. S3).

To better characterize the vaquita's long-term demographic history, we used the distribution of allele frequencies to perform model-based demographic inference. Overall, we found good fit for a two-epoch model in which the vaquita effective population size (N_e) declined from 4485 to 2807 individuals ~2162 generations ago (~25.7 thousand years ago) (*5*) (Fig. 1E, figs. S4 and S5, and tables S2 to S4). Thus, vaquitas have persisted at relatively small population sizes for at least tens of thousands of years, which has resulted in uniformly low genome-wide diversity that is among the lowest documented in any species to date (*12*). Here, we use the phrase "long-term small population size" to mean N_e on the order of a few thousand individuals over thousands of generations, as opposed to $N_e \leq 100$, as in some other contexts (*14*, *15*).

A predicted consequence of long-term small population size is the reduced efficacy of purifying selection against weakly deleterious alleles with selection coefficients $\ll 1/(2 \times N_e)$ (*14*, *15*). Such alleles can drift to high frequencies and become fixed, potentially contributing to reduced fitness. To investigate this, we compared the burden of putatively deleterious protein-coding variants in vaquitas with those in 11 other cetacean species (table S5 and fig. S6). Specifically, we focused on nonsynonymous mutations at sites under strong evolutionary constraint (*16*) and loss-of-function (LOF) mutations that are predicted to disrupt gene function. We used the ratio of deleterious to synonymous variants as a proxy for the efficacy of purifying selection (*5*) and used genome-wide heterozygosity as a proxy for N_e (Fig. 2, A and B, and fig. S7). The ratio of deleterious variants is significantly negatively correlated with N_e [phylogenetic generalized least squares (PGLS) regression, $p_{del} = 1.32 \times 10^{-2}$, $p_{LOF} = 7.88 \times 10^{-3}$), consistent with expectation. Among all species in our study, vaquitas have the highest proportional burden of deleterious alleles. Compared with the species with the next lowest diversity (the orca, *Orcinus orca*), ratios for deleterious and LOF mutations in vaquitas are 1.14× and 1.23× higher, respectively. Furthermore, we demonstrate using simulations that this elevated ratio is minimally affected by the vaquita's recent population decline and is instead attributable to its historical population size (fig. S9) (*5*). Similar trends exist for homozygous deleterious mutations, which include variants that may be fixed in the species (fig. S8). Thus, elevated ratios of deleterious-to-neutral variation among polymorphisms (heterozygotes) and substitutions (homozygotes) in vaquitas are consistent with an accumulation of weakly deleterious alleles under long-term small population size. However, despite this elevated burden of weakly deleterious variants, the remaining vaquita individuals appear healthy and are actively reproducing (*17*, *18*), which suggests that the species' fitness has not been severely compromised.

A larger concern for vaquita recovery is future fitness declines resulting from inbreeding depression given the inevitability of inbreeding in any recovery scenario. However, the risk of inbreeding depression (inbreeding load) is predicted to be reduced in species with long-term small population size because (i) increased homozygosity exposes recessive strongly deleterious alleles to selection more frequently and (ii) drift decreases the absolute number of segregating recessive deleterious variants (*19*, *20*). To assess the potential for future inbreeding depression in vaquitas relative to other cetaceans, we quantified the total number of heterozygous deleterious alleles per genome, which reflect alleles that could contribute to

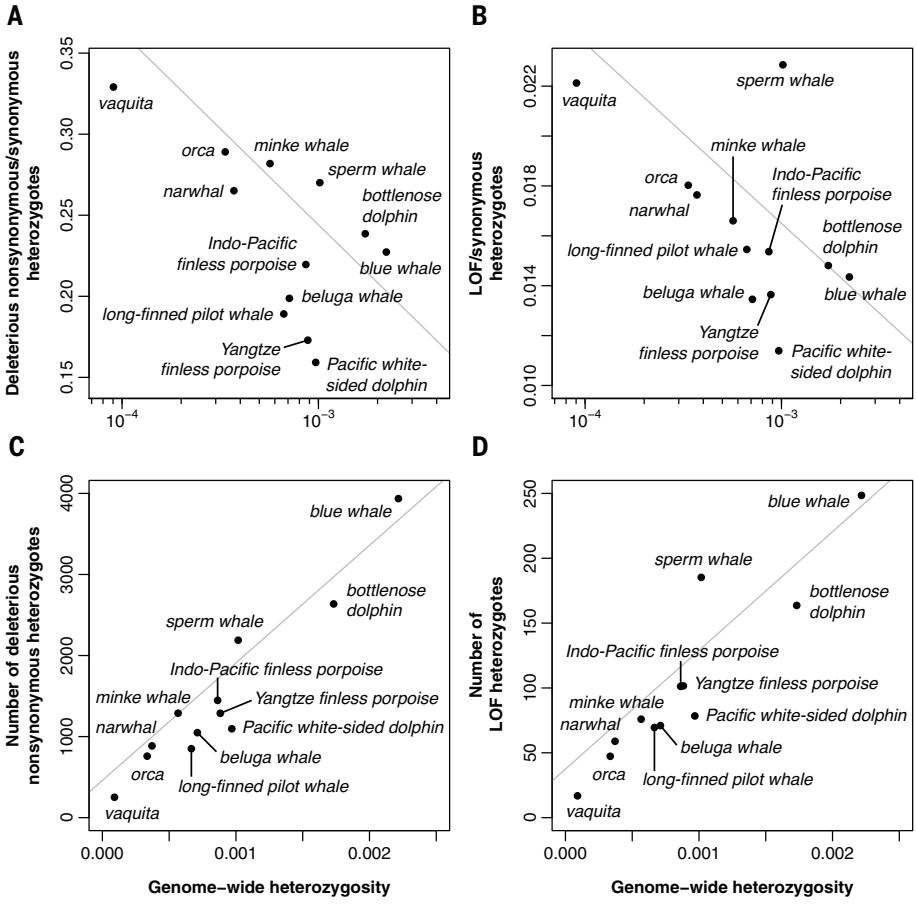


Fig. 2. Deleterious variation in vaquitas and other cetaceans. (**A** and **B**) Ratios of deleterious nonsynonymous (A) and LOF (B) heterozygotes to synonymous heterozygotes are significantly negatively correlated with genome-wide heterozygosity (per base pair, log-scaled). (**C** and **D**) Total numbers of deleterious nonsynonymous (C) and LOF (D) heterozygotes per genome are significantly positively correlated with genome-wide heterozygosity (per base pair). Gray lines show phylogeny-corrected regressions [excluding the Indo-Pacific finless porpoise (*5*)].

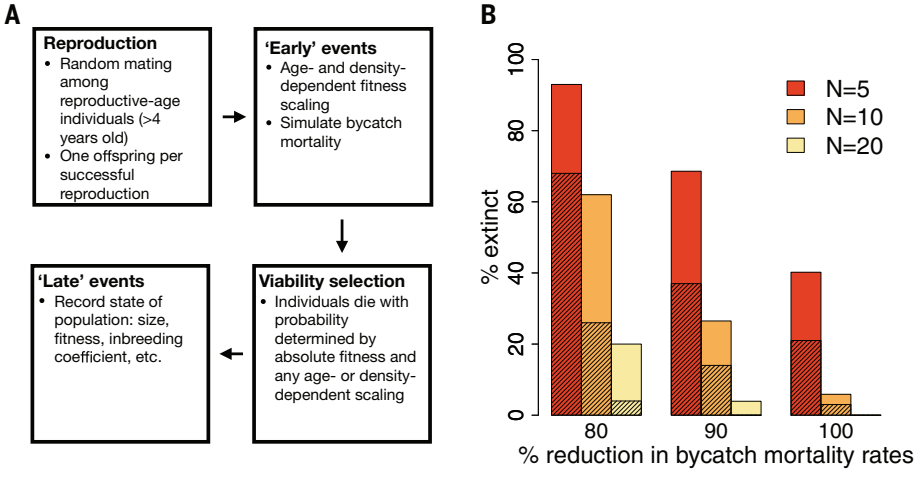


Fig. 3. Model schematic and extinction rates under various simulation parameters. (**A**) Diagram of events that occur during 1 year in our SLiM simulation model. (**B**) Percent of replicates going extinct over the next 50 years under varying recovery parameters. Shading indicates extinction rates when only neutral mutations are simulated, and N values represent the threshold population sizes.

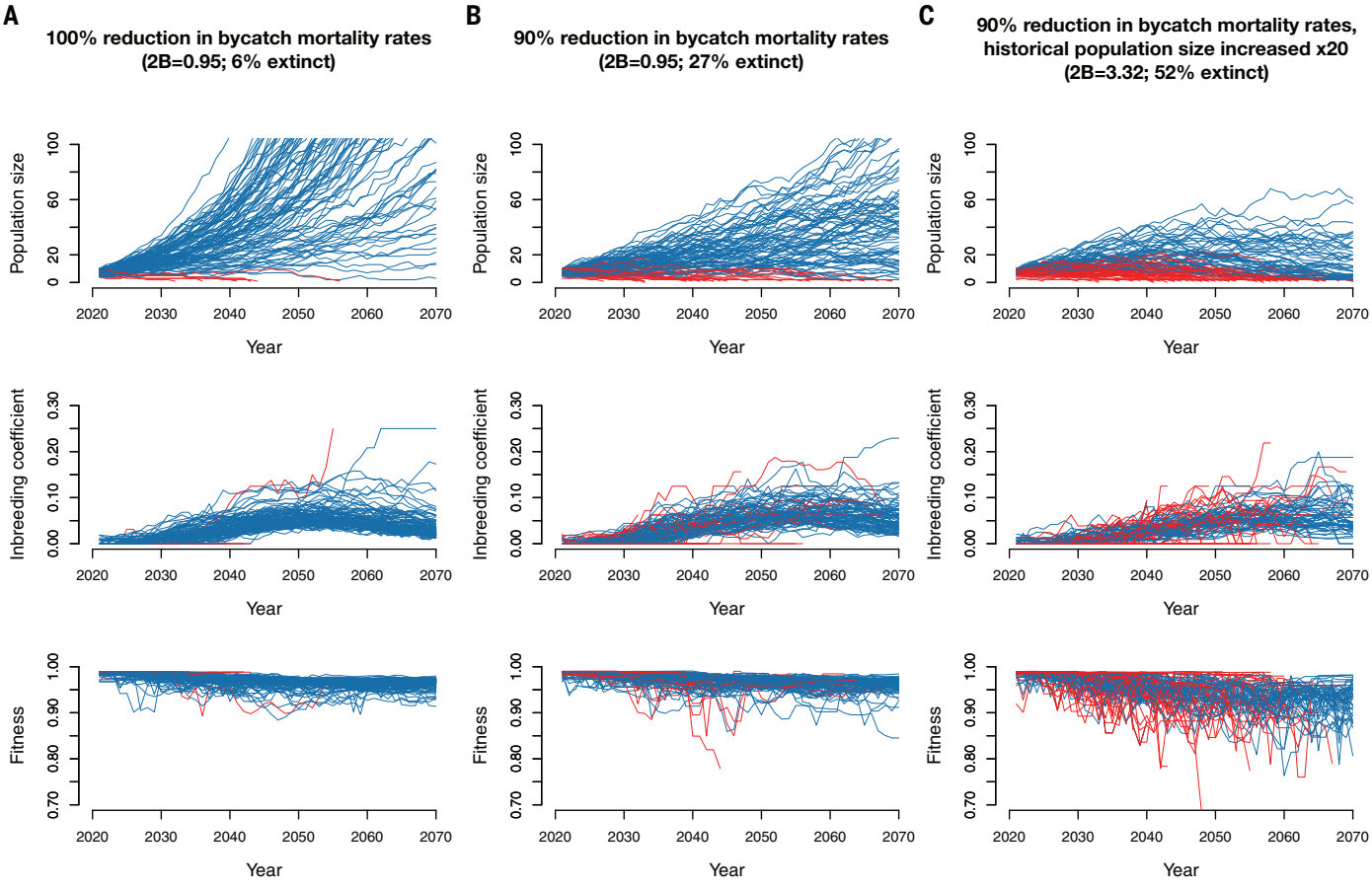


Fig. 4. Simulation trajectories under various recovery scenarios. (**A**) Simulation trajectories under empirically inferred historical demographic parameters assuming a reduction in bycatch mortality of 100%. (**B**) Simulation trajectories with bycatch mortality rate decreased by only 90%. (**C**) Simulation trajectories with historical population size increased by 20× and assuming a decrease in bycatch mortality of 90%. For all simulations shown here, we assumed a population size threshold of 10 individuals. Replicates that went extinct are colored red, and replicates that persisted are colored blue.

inbreeding depression when made homozygous through inbreeding. We found that the total number of heterozygous putatively deleterious alleles per genome is positively correlated with genome-wide diversity (PGLS p_{del} = 5.57×10^{-6}, $p_{LOF} = 1.91 \times 10^{-5}$) (Fig. 2, C and D). Among all cetaceans in our study, vaquitas harbor the fewest deleterious heterozygotes per genome. Compared with the orca, vaquitas have 0.33× and 0.36× the number of deleterious and LOF heterozygotes, respectively. Similar trends are evident in all mutation classes, including conserved noncoding regions (fig. S10). Thus, although vaquitas have an elevated proportion of deleterious relative to neutral variants (Fig. 2, A and B, and fig. S8), they nevertheless have a low absolute number of segregating deleterious variants (Fig. 2, C and D), which implies a low inbreeding load.

To model potential recovery scenarios for the vaquita, we combined our genomic results with information about vaquita life history to parameterize stochastic, individual-based simulations using SLiM3 (*5*, *21*) (Fig. 3A and fig. S11). These simulations were designed to model vaquita protein-coding regions, incorporating neutral mutations and (partially) recessive deleterious mutations, the latter of which are thought to underlie inbreeding depression (*3*, *22*). We used our genomic dataset to estimate a vaquita mutation rate (fig. S12) as well as a distribution of selection coefficients for new mutations (fig. S13) and assumed an inverse relationship between dominance and selection coefficients (*5*). Notably, our model allows for deleterious mutations to drift to fixation and affect fitness (figs. S14 to S16) (*5*). We used our demographic model (Fig. 1E) to simulate the historical vaquita population (figs. S17 and S18) and then initiated a bottleneck by introducing stochastic bycatch mortality at a rate calibrated to the empirical rate of recent decline as of 2018 (Fig. 1A and fig. S19) (*5*). Finally, we allowed for recovery by reducing the bycatch mortality rate after the population reached a threshold population size of 10 or fewer individuals, on the basis of the current estimated population size.

We first used this model to examine the impact of varying levels of bycatch mortality on extinction risk over the next 50 years. We estimate a high probability of recovery if bycatch mortality ceases entirely, with only 6% of simulation replicates going extinct (Fig. 3B and Fig. 4A). Additionally, simulated populations that persist exhibit substantial growth, with a mean population size in 2070 of 298.7 individuals (SD = 218.2; Fig. 4A). However, if bycatch mortality rates are decreased by just 90%, extinction rates increase to 27% (Fig. 3B and Fig. 4B), with more-limited recovery in population sizes (mean of 49.2 individuals in 2070, SD = 34.4; Fig. 4B). Finally, if bycatch mortality rates are decreased by just 80%, extinction occurs in 62% of simulation replicates. Thus, recovery potential critically depends on reducing bycatch mortality rates, with even moderate levels of bycatch resulting in a high likelihood of extinction.

Next, we examined the importance of the threshold population size, given uncertainty in the 2018 estimate of 10 individuals (*4*). As expected, extinction rates decrease when assuming a threshold population size of 20 and increase when assuming a threshold

population size of five (Fig. 3B). These results emphasize that the number of remaining vaquita individuals is also a critical factor underlying extinction risk.

To quantify the inbreeding load in our model, we estimated the number of diploid lethal equivalents ($2B$), which characterizes the rate at which fitness is lost with increasing levels of inbreeding (*2*, *23*). Typically, inbreeding load is quantified by comparing estimates of individual fitness and inbreeding in natural populations (*2*, *24*); however, such data do not exist for most species, including the vaquita. Under our simulation parameters, we estimate an inbreeding load of $2B = 0.95$ in vaquitas (table S6), which is substantially lower than the median empirical estimate for mammals of 3.1 (*24*), likely because of the vaquita's relatively small historical N_e. Nevertheless, simulations that exclude deleterious mutations result in a substantially lower extinction rate (Fig. 3B), which confirms that inbreeding depression affects recovery potential in our model.

To further explore how the inbreeding load in our model depends on historical demography, we ran simulations with the historical N_e increased by 20×. We found an increased extinction rate of 52%, compared with 27% with our empirical population size parameters, with minimal recovery for replicates that persisted (mean of 16.2 individuals in 2070, SD = 14.5; Fig. 4C). Additionally, with this larger historical N_e, we observe a greatly increased inbreeding load of $2B = 3.32$ (fig. S20 and table S6). These findings further demonstrate the importance of the vaquita's natural rarity as a factor underlying their low inbreeding load and increased potential for recovery.

Given the uncertainty in many of our model parameters, we conducted sensitivity analyses varying the calving interval, mutation rate, distribution of dominance and selection coefficients, and target size for deleterious mutations (*5*). Although these factors influence extinction probabilities, recovery remains the likely outcome (>50% probability) in nearly all cases when assuming a threshold population size of 10 and a 90% reduction of bycatch mortality (fig. S21 and table S6). Two notable exceptions to this are for models with a higher mutation rate, where we observed a 55% extinction rate compared with 27% in our base model, and for models with a decreased calving interval, where we also observed a 55% extinction rate (fig. S21 and table S6). Thus, although uncertainty exists in our projections, the overall conclusion that recovery is possible if bycatch is greatly reduced remains robust to our model assumptions. Finally, we note that our simulations do not consider factors such as reduced adaptive potential or increased susceptibility to disease caused by low genetic variability, which may affect future persistence. Vaquitas have survived with low diversity for tens of thousands of years and have endured environmental changes in the past (*12*), which suggests that these factors alone do not doom the species to extinction. Conceivably, low diversity in the vaquita may limit the species' capacity to adapt to increasing global change over the long term, but this risk is challenging to quantify and should not preclude recovery efforts in the short term.

Our results suggest that there is a high potential for vaquita recovery in the absence of gillnet mortality, refuting the view that the species is doomed to extinction by genetic factors. Our approach leverages genomic data and methodology to forecast population viability and extinction risk, which enables a more nuanced assessment of the threat of genetic factors to persistence. The key aspect of the vaquita that our analysis reveals is that its historical population size was large enough to prevent the fixation of all but weakly deleterious alleles and small enough to reduce the inbreeding load from recessive, strongly deleterious mutations. Numerous other examples of species rebounding from bottlenecks of a similar magnitude to that of the vaquita have been documented (*1*). For example, many parallels exist between the vaquita and Channel Island foxes, which similarly have exceptionally low genetic diversity, yet were able to rebound from severe recent bottlenecks without apparent signs of inbreeding depression (*25*). Together, these examples challenge the assumption that populations that have experienced catastrophic declines are genetically doomed and provide hope for the recovery of endangered species that are naturally rare. Finally, our analysis demonstrates the potential for genomics-informed population viability modeling, which may have widespread applications given the increasing feasibility of genomic sequencing for nonmodel species amid a worsening extinction crisis (*26*).

REFERENCES AND NOTES

1. D. A. Wiedenfeld *et al*., *Conserv. Biol.* **35**, 1388–1395 (2021).
2. L. F. Keller, D. M. Waller, *Trends Ecol. Evol.* **17**, 230–241 (2002).
3. D. Charlesworth, J. H. Willis, *Nat. Rev. Genet.* **10**, 783–796 (2009).
4. A. M. Jaramillo-Legorreta *et al*., *R. Soc. Open Sci.* **6**, 190598 (2019).
5. The materials and methods and supplementary text are available as supplementary materials.
6. M. A. Cisneros-Mata, J. A. Delgado, D. Rodríguez-Félix, *Rev. Biol. Trop.* **69**, 588–600 (2021).
7. B. L. Taylor, L. Rojas-Bracho, *Mar. Mamm. Sci.* **15**, 1004–1028 (1999).
8. C. Sonne, P. Diaz-Jaimes, D. H. Adams, *Science* **373**, 863–864 (2021).
9. B. W. Brook *et al*., *Nature* **404**, 385–387 (2000).
10. F. W. Allendorf, P. A. Hohenlohe, G. Luikart, *Nat. Rev. Genet.* **11**, 697–709 (2010).
11. H. A. Lewin *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **115**, 4325–4333 (2018).
12. P. A. Morin *et al*., *Mol. Ecol. Resour.* **21**, 1008–1020 (2021).
13. B. L. Taylor, S. J. Chivers, J. Larese, W. F. Perrin, "Generation length and percent mature estimates for IUCN assessments of cetaceans" (Southwest Fisheries Science Center, Administrative Report LJ-07-01, 2007).
14. M. Lynch, I. J. Conery, R. Burger, *Am. Nat.* **146**, 489–518 (1995).
15. M. Kimura, T. Maruyama, J. F. Crow, *Genetics* **48**, 1303–1312 (1963).
16. P. C. Ng, S. Henikoff, *Genome Res.* **11**, 863–874 (2001).
17. B. L. Taylor *et al*., *Mar. Mamm. Sci.* **35**, 1603–1612 (2019).
18. F. Gulland *et al*., *Vet. Rec.* **187**, e51 (2020).
19. C. C. Kyriazis, R. K. Wayne, K. E. Lohmueller, *Evol. Lett.* **5**, 33–47 (2021).
20. S. Glémin, *Evolution* **57**, 2678–2687 (2003).
21. B. C. Haller, P. W. Messer, *Mol. Biol. Evol.* **36**, 632–637 (2019).
22. A. R. McCune *et al*., *Science* **296**, 2398–2401 (2002).
23. N. E. Morton, J. F. Crow, H. J. Muller, *Proc. Natl. Acad. Sci. U.S.A.* **42**, 855–863 (1956).
24. K. Ralls, J. D. Ballou, A. Templeton, *Conserv. Biol.* **2**, 185–193 (1988).
25. J. A. Robinson, C. Brown, B. Y. Kim, K. E. Lohmueller, R. K. Wayne, *Curr. Biol.* **28**, 3487–3494.e4 (2018).
26. G. Ceballos, P. R. Ehrlich, P. H. Raven, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 13596–13602 (2020).
27. J. Robinson, S. Nigenda, A. Beichman, jarobin/vaquitagenomics2022: v1, version v1, Zenodo (2022); https://doi.org/10.5281/zenodo.6303135.
28. C. C. Kyriazis, ckyriazis/vaquita_simulations, *Zenodo* (2022); https://doi.org/10.5281/zenodo.6308771.

ACKNOWLEDGMENTS

We thank the CanSeq150 project for use of the long-finned pilot whale and Pacific white-sided dolphin genomes. Y. Bukhman generously provided early access to the blue whale genome. We thank J. Mah, P. Nuñez, and M. Lin for providing scripts; B. Haller for assistance with simulations; and K. Harris for advice on the inference of selection coefficients. We thank the Southwest Fisheries Science Center's Marine Mammal and Sea Turtle Research Collection for use of archival vaquita tissue samples. All samples were imported to the US under appropriate CITES and US Marine Mammal Protection Act permits. **Funding:** We thank F. Gulland, The Marine Mammal Center, and NOAA Fisheries for funding genome resequencing. C.C.K. and K.E.L. were supported by National Institutes of Health (NIH) grant R35GM119856 (to K.E.L.). A.C.B. was supported by the Biological Mechanisms of Healthy Aging Training Program, NIH T32AG066574. S.F.N.-M. was supported by the Mexican National Council for Science and Technology (CONACYT) postdoctoral fellowship 724094 and the Mexican Secretariat of Agriculture and Rural Development postdoctoral fellowship. **Author contributions:** P.A.M., B.L.T., K.E.L., J.A.R., and C.C.K. designed the study. P.A.M., M.C.F., L.R.-B., and B.L.T. obtained funding. P.A.M., B.L.T., and L.R.-B. obtained samples. K.M.R. performed DNA extractions and library preparations. A.C.B., S.F.N.-M., J.A.R., and C.C.K. performed analyses. J.A.R. and C.C.K. wrote the manuscript with input from all authors. P.A.M., B.L.T., K.E.L., and R.K.W. supervised the work. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** Vaquita raw sequence reads have been deposited in the Sequence Read Archive (SRA) under BioProject PRJNA751981 (see table S1 for details). Accession information for publicly available cetacean genomes is provided in table S5. Scripts used for sequence data processing and analysis (*27*) and scripts for simulations (*28*) are available on Zenodo.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm1742
Materials and Methods
Supplementary Text
Figs. S1 to S21
Tables S1 to S6
References (*29*–*106*)
MDAR Reproducibility Checklist

Submitted 31 August 2021; accepted 2 March 2022
10.1126/science.abm1742

ICE SHEETS

A dynamic saline groundwater system mapped beneath an Antarctic ice stream

Chloe D. Gustafson[1,2]*, Kerry Key[1], Matthew R. Siegfried[3], J. Paul Winberry[4], Helen A. Fricker[2], Ryan A. Venturelli[5], Alexander B. Michaud[6]

Antarctica's fast-flowing ice streams drain the ice sheet, with their velocity modulated by subglacial water systems. Current knowledge of these water systems is limited to the shallow portions near the ice-bed interface, but hypothesized deeper groundwater could also influence ice streaming. Here, we use magnetotelluric and passive seismic data from Whillans Ice Stream, West Antarctica, to provide the first observations of deep sub–ice stream groundwater. Our data reveal a volume of groundwater within a >1-kilometer-thick sedimentary basin that is more than an order of magnitude larger than the known subglacial system. A vertical salinity gradient indicates exchange between paleo seawater at depth and contemporary basal meltwater above. Our results provide new constraints for subglacial water systems that affect ice streaming and subglacial biogeochemical processes.

The ice flux from the Antarctic Ice Sheet's interior to its margins is largely governed by the behavior of fast-flowing ice streams [e.g., (*1*)], where the availability of subglacial water plays a fundamental role in regulating ice flow (*2*). Given the inaccessibility of the Antarctic substrate, physical constraints on the sub–ice stream water systems are based on a few geophysical (i.e., active seismic, radar sounding) surveys [e.g., (*3*, *4*)] and subglacial drilling efforts [e.g., (*5*)]. These observations revealed a "shallow" sub–ice stream hydrologic system, which we define here as water at the ice-bed interface (in the form of films, channels, and lakes) and water within a relatively thin (≲10 m) deformable layer of unsorted sediments, generally considered glacial till (*5*). Below these shallow hydrologic systems, Antarctic ice streams are expected to be underlain by sedimentary basins hundreds to thousands of meters thick that have the capacity to host substantial groundwater (*6*), constituting "deep" subglacial hydrologic systems, which we define here as groundwater below the glacial till. These deep groundwater systems have the potential to provide water to or remove water from the ice base and, therefore, may play a role in modulating ice flow (*6*). However, current Antarctic subglacial water models only consider the shallow hydrologic systems, partly because the geophysical tools typically used to map deep sub–ice stream sediments (i.e., active and passive seismics, gravity, and magnetics) (*7*–*9*) do not provide constraints on deep groundwater properties or connectivity between the shallow and deep hydrologic systems. Here, we use magnetotelluric (MT) (*10*) and passive seismic data (*11*, *12*) to provide the first in situ measurements of a deep groundwater system within a >1-km-thick sedimentary basin beneath the West Antarctic Ice Sheet. We infer that this basin contains more than an order of magnitude more water than the shallow hydrologic systems typically considered in subglacial water

[1]Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY, USA. [2]Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA, USA. [3]Hydrologic Science and Engineering Program, Department of Geophysics, Colorado School of Mines, Golden, CO, USA. [4]Department of Geological Sciences, Central Washington University, Ellensburg, WA, USA. [5]Department of Geology and Geological Engineering, Colorado School of Mines, Golden, CO, USA. [6]Bigelow Laboratory for Ocean Sciences, East Boothbay, ME, USA.
*Corresponding author. Email: cgustafson@ucsd.edu

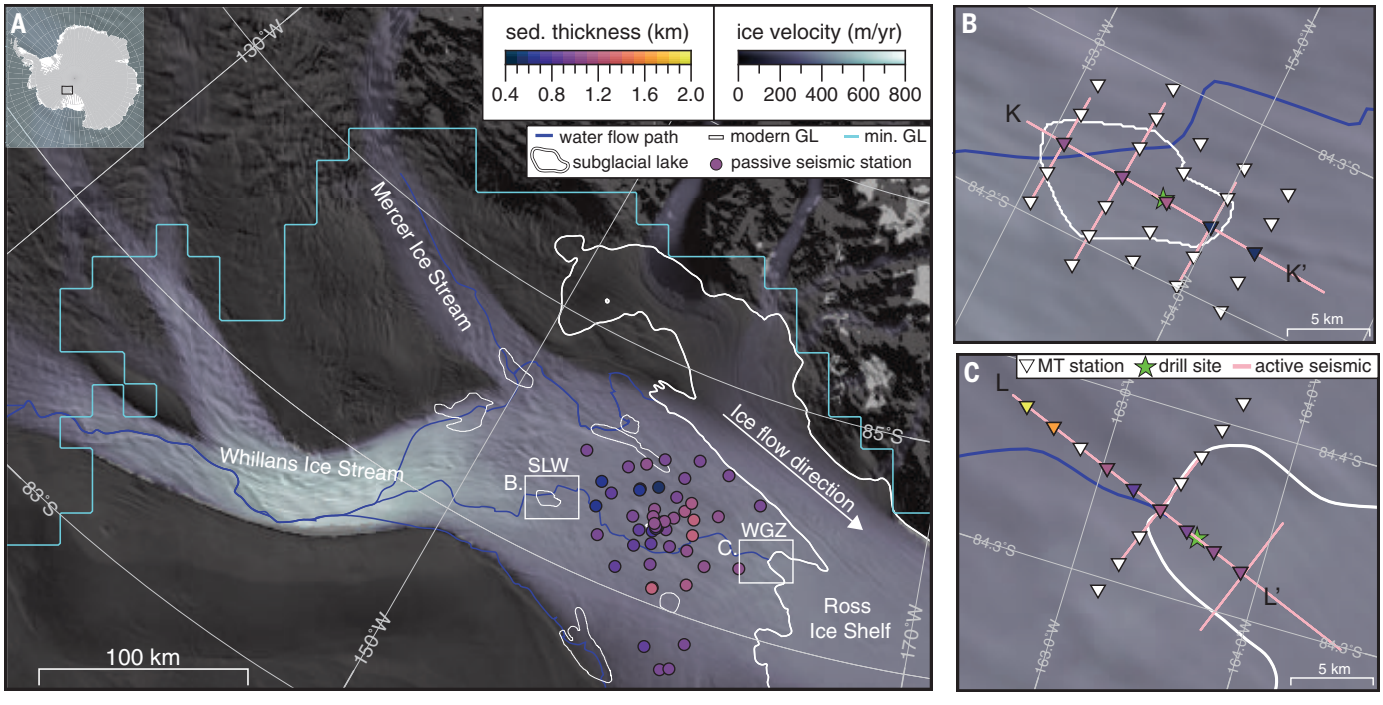


Fig. 1. Regional map of passive seismic and MT surveys on Whillans Ice Stream, West Antarctica. (**A**) WIS ice velocity (*44*) plotted over a satellite image mosaic (*45*) with active subglacial lakes (*20*) and predicted basal water flow paths (*42*) that deliver water across the modern grounding line (GL) (*46*). The estimated most-landward grounding line position after the Last Glacial Maximum (*34*) is shown in cyan (minimum GL, labeled as "min. GL"). Colored circles indicate passive seismic receiver locations and corresponding sediment thickness estimates. (**B** and **C**) MT stations at (B) SLW and (C) WGZ are aligned with previous active-source seismic profiles (*21*, *22*) and sites of direct subglacial drilling access (*23*, *24*). Colored triangles indicate sediment thickness estimates (*24*). 2D inversion results of MT stations K to K' and L to L' are shown in Fig. 2.

models (*13*). The groundwater salinity increases with depth, indicating that the deep and shallow subglacial water systems are hydrologically connected and that Antarctic groundwater may influence ice stream behavior and subglacial biogeochemical processes.

Previous studies have recognized the potential presence of groundwater beneath Antarctic ice streams [e.g., (*6*)]; however, its detection has been challenging. Electromagnetic (EM) methods, which are frequently used for terrestrial groundwater studies (*14*), are also a feasible technique for detecting deep subglacial groundwater (*15*). Variations in resistivity can be exploited to delineate high-resistivity glacier ice (~4×10^4 to 1×10^8 ohm·m) from low-resistivity sediments (~1 to 100 ohm·m) and moderately resistive bedrock (~100 to 1000 ohm·m) (*15*). Furthermore, resistivity variations within groundwater-bearing sediments can provide constraints on water properties because saltier porewater yields a bulk resistivity that is an order of magnitude lower than that of freshwater (*15*). Recent surveys in the McMurdo Dry Valleys have demonstrated the effectiveness of high-frequency airborne EM soundings for mapping groundwater within the top few hundred meters of subglacial environments [e.g., (*16*)] but under only <350 m of ice. Sensing deeper groundwater within sedimentary basins below thicker ice requires ground-based EM methods, such as MT (*15*), a passive EM method that uses natural low-frequency EM fields to image subsurface resistivity at depths spanning from the shallow crust to the upper mantle. MT methods have been widely applied in continental and oceanic settings for constraining subsurface fluid and lithologic distributions (*17*) yet, in Antarctica, have only been used to examine the deep crust and lithosphere (*18*).

We collected MT data (fig. S1) on the downstream outlet of Whillans Ice Stream (WIS), located along Siple Coast (which we use here to collectively describe the Gould, Siple, and Shirase coasts) of West Antarctica (Fig. 1A), to investigate deep groundwater within a sedimentary basin under ~800 m of ice. Numerical simulations have estimated that groundwater considerably contributes to Siple Coast ice streaming, but this modeling only included shallow water volumes and fluxes within the till because there were no observations of deep groundwater to constrain their modeling (*19*). Our MT surveys focused on two hydrologically connected regions that are part of a dynamic subglacial water system (*20*): Whillans Subglacial Lake (SLW) (Fig. 1B) and its hypothesized downstream outlet to the ocean, Whillans Grounding Zone (WGZ) (Fig. 1C). We supplemented our focused MT surveys with a passive seismic survey that was sparser but sampled a wider region, extending across WIS (colored circles in Fig. 1A). At both MT survey locations, the shallow hydrology has been constrained by geophysical surveying (*4*, *21*, *22*) and direct subglacial sampling (*23*, *24*). Although these studies did not image or sample the deeper groundwater, porewater samples from a shallow (0.38 m) sediment core below SLW showed an increasing contribution of seawater-sourced Cl^-, indicating a deeper reservoir of seawater (*25*). The active-source seismic studies (pink lines in Fig. 1, B and C) also revealed that each location is underlain by a substantial thickness

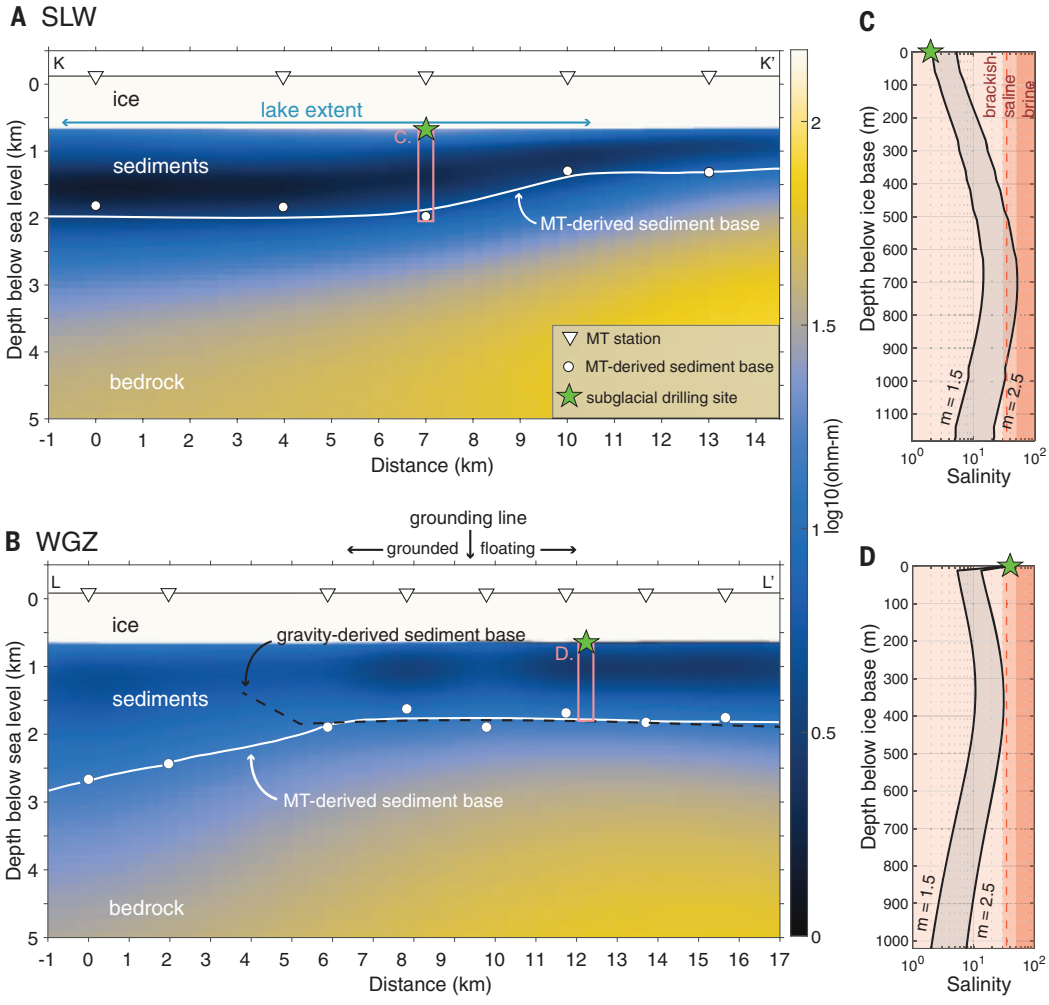


Fig. 2. Resistivity models and inferred groundwater salinity. (**A** and **B**) Electrical resistivity models of (A) SLW and (B) WGZ obtained by 2D regularized inversion (*27*). White circles show the estimated sediment thickness obtained from a 1D Bayesian inversion of each station's data [fig. S4 and (*27*)]. The WGZ model additionally includes a black dashed line noting the base of the sediment layer as interpreted by ground-based gravity surveying (*28*). The deviation between the MT- and gravity-derived sediment base around the 4-km position is likely an edge effect in the gravity inversion. (**C** and **D**) Vertical water salinity profile estimates (gray-shaded region) estimated by applying Archie's law (*27*) to vertical profiles extracted from the 2D resistivity models. Red shading denotes the salinity characterizations using the unitless practical salinity scale [brackish (0.1 to 30), saline (30 to 50), and brine (>50)], whereas the red dashed line denotes seawater salinity (35). The green stars in (A) and (B) note the location of subglacial sediment sampling. In (C), the green star marks the maximum porewater salinity from subglacial sampling of the top 0.38 m of sediments at the same location (*25*), and in (D), the green star indicates the seawater salinity observed in the ocean cavity (*24*).

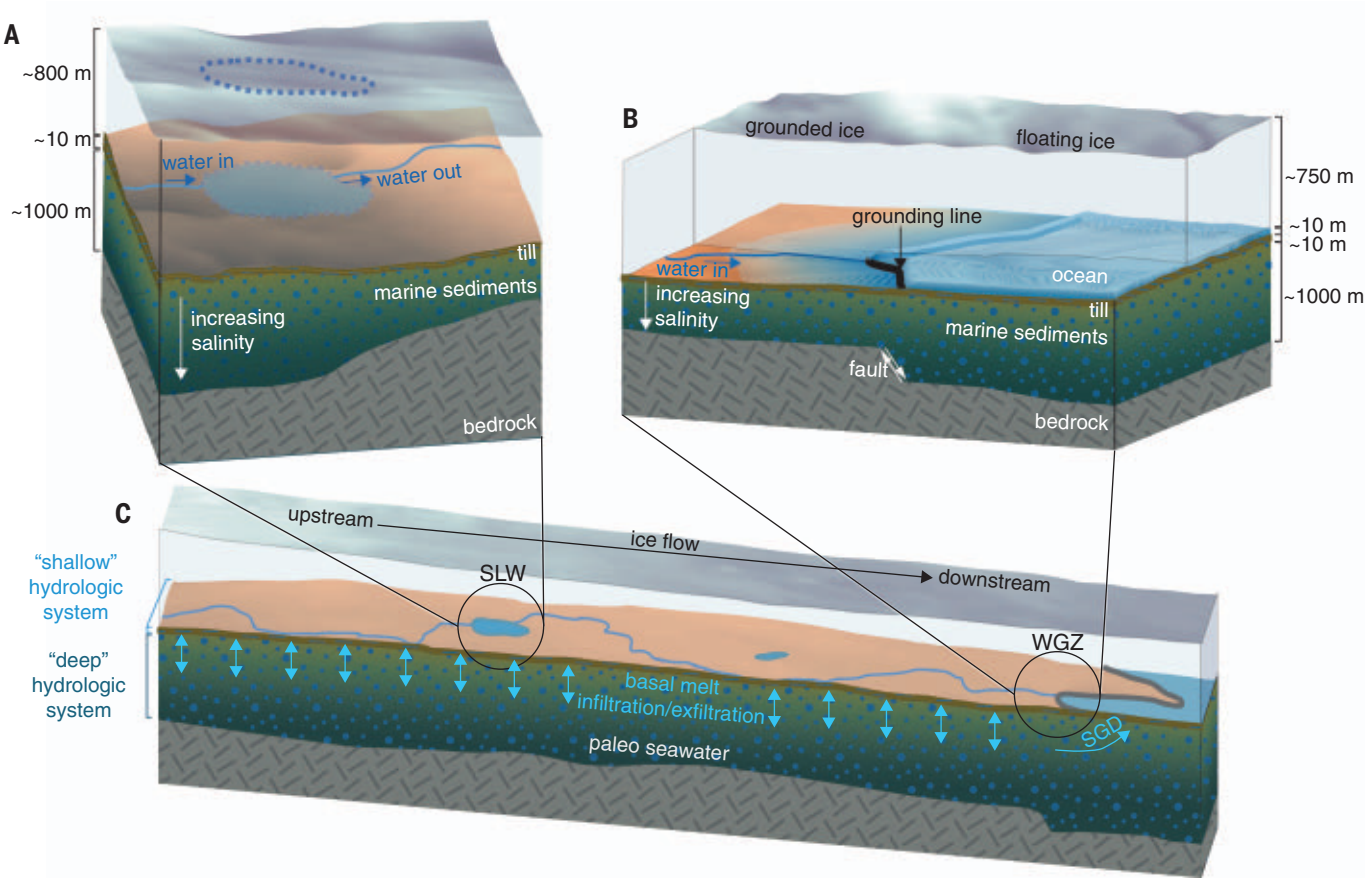


Fig. 3. Conceptual models of WIS groundwater and water routing. (**A** to **C**) Thick sediments that contain groundwater that increases in salinity with depth at both (A) SLW and (B) WGZ are persistent throughout (C) WIS. We attribute the vertical salinity gradient to paleo seawater mixing with contemporary basal meltwater. Low-salinity groundwater in sub–ocean cavity sediments (Fig. 2D) indicates that groundwater flows laterally across the grounding line and enters the ocean as SGD.

of stratified sediments (>100 m) (*22*, *26*) but were not able to conclusively map depth to bedrock (i.e., total sediment thickness) or constrain the deep hydrology.

We determined sediment thickness beneath WIS using passive seismic and MT observations. Our regional passive seismic sediment thickness estimates obtained by receiver function analysis (*27*) (figs. S2 and S3) range from 0.6 to 1.3 km (± 0.2 km) (Fig. 1A), which are consistent with MT-derived sediment thickness estimates that we calculated using one-dimensional (1D) Bayesian inversion methods (*27*) (fig. S4) at SLW (0.5 to 1.2 km ± 0.1 km; Fig. 1B) and at WGZ (0.9 to 1.9 km ± 0.1 km; Fig. 1C). These estimates are in overall agreement with those derived from airborne surveying [e.g., (*7*)] and ground-based gravity surveying (*28*). The Bayesian sediment thickness estimates are also compatible with the structure of the 2D resistivity models that we obtained using regularized inversion for linear profiles of MT stations (*27*), which revealed a low-resistivity layer consistent with water-saturated sediments overlying a more resistive layer consistent with bedrock (Fig. 2, A and B, and figs. S5 and S6).

Information about groundwater contained within the sediments comes from examining resistivity within the sediment layer in our 2D models (blue shading above the white line in Fig. 2, A and B). We used the empirical Archie's law to convert resistivity to groundwater salinity for vertical profiles in our model that are colocated with past subglacial drilling sites at SLW (Fig. 2C) and WGZ (Fig. 2D) (*24*). We assumed a range of Archie's exponents that are appropriate for sedimentary rocks (m = 1.5 to 2.5), a sediment compaction model consistent for porosity observations from the Ross Sea (*29*), and a 50°C/km temperature gradient consistent with heat-flux observations from WIS (*30*). Salinity at the top of the sediment column is brackish (defined here as 0.1 to 30 on the unitless practical salinity scale) for both SLW (2 to 5) and WGZ (5 to 13). At SLW, this range corresponds well with direct porewater salinity measurements [green star in Fig. 2C, and (*25*)]. At WGZ, we constrain our model to include a seawater layer (*24*) that matches observed seawater conductivity [green star in Fig. 2D and (*24*)]. Groundwater salinity increases with depth, to 700 m at SLW and 400 m at WGZ, with the upper bounds at both locations reaching values >30, that is, close to that of seawater (~35). At both locations, the salinity appears to decrease below these depths. However, because MT data cannot resolve a sharp increase in resistivity that would result from low-resistivity salty groundwater in direct contact with high-resistivity bedrock (*14*), we propose that the saltiest groundwater is located at the bottom of the sediment column (Fig. 3, A and B) and the apparent freshening is a result of smoothing in our regularized inversion.

Given the similarity in groundwater salinity and general resistivity structure between SLW and WGZ, we infer that the thick sediments sampled by our passive seismic stations are also saturated with groundwater that increases in salinity with depth (Fig. 3C). If we were to extract this water from the sediments, the water column's equivalent thickness would range from 220 to 820 m, assuming the same porosity model from our Archie's law calculations integrated over the thickness of the sediments estimated from MT and passive seismic data (0.4 to 1.9 km) (*27*). This demonstrates that the volume of deep groundwater is

at least an order of magnitude greater than the water in shallow sub–ice stream hydrologic systems [~2 to 15 m for Siple Coast ice stream subglacial lakes (*31*, *32*) and ~3-m equivalent water depth for the till (*13*)].

The upward freshening of groundwater provides information about the glacial and hydrologic processes that led to the development of the groundwater system, as well as the extent of the system. We propose that the salty porewater within the deepest sediments is of marine origin, which is consistent with geochemistry from the shallow subglacial system at SLW (*25*). The most recent marine incursion at WIS occurred in the mid-Holocene (*33*) following the Last Glacial Maximum deglaciation, when the grounding line retreated inland of its modern position, upstream of SLW (*34*) (cyan line in Figs. 1A and 4). We propose that this resulted in seawater rapidly convecting downward to the base of the preexisting sediments [fig. S7 and (*27*, *35*)] beneath WIS and other Siple Coast ice streams (cyan hatching in Fig. 4) that experienced the same marine incursion (*33*, *34*). Some deep salty water may also be left over from when the sediments were initially deposited under open-ocean conditions when West Antarctica was ice free (*36*); this older water possibly extends upstream to the onset of ice streaming (blue hatching in Fig. 4). The emplacement of lower salinity water that we observe at the top of the sediments was driven by subsequent grounding line readvance and ice sheet regrowth, which forced fresh basal meltwater into the sediments through ice-sediment hydromechanical feedbacks [e.g., (*37*)].

Infiltration of basal meltwater into the top few hundred meters of sediments demonstrates that the deep and shallow sub–ice stream hydrologic systems are physically connected. This connection has been suggested theoretically using hydromechanical models of paleo (*37*) and extant (*38*) ice sheets but has not previously been proven through observations in Antarctica. Although our observed salinity gradients are only indicative of shallow subglacial water infiltrating into the deep hydrologic environment, deep groundwater can also ascend into the shallow system. Because ice thickening promotes water flow into the sediments and ice thinning allows for water to exfiltrate from the sediments and enter the shallow hydrologic environment (*37*), the direction of water flow and the spatiotemporal scales over which flow occurs is dynamically coupled to ice behavior.

The confirmation of the existence of deep groundwater dynamics has transformed our understanding of ice stream behavior and will force modification of subglacial water models. A connection between the deep and shallow subglacial water systems suggests that upward groundwater flow is another potential source of water and heat (*38*), which can enhance basal melting at the ice base. Conversely, downward groundwater flow removes both water and heat from the ice base (*38*), promoting freezing conditions that slow ice flow. The role of groundwater flow may be further complicated by the transport of solutes between the deep and shallow hydrologic systems, which can modify the in situ basal melting point. Given the large groundwater volume that we observe, we propose that future Antarctic sub–ice stream water models incorporate both deep and shallow water systems to determine the importance of groundwater to ice stream dynamics.

The presence and flow of deep groundwater also has implications for subglacial ecosystems and biogeochemical cycles. Our observations of deep paleo seawater suggest that Siple Coast sedimentary basins are marine in origin, and biogeochemical cycles common to deeply buried marine sediments (i.e., methanogenesis) are likely operating and may be responsible for carbon transformation at depth (*39*, *40*). Upward groundwater fluxes may transport the products produced by microbial communities (e.g., dissolved organic or inorganic carbon) to the shallow subglacial systems, where it fuels subglacial ecosystems (*23*, *41*) and can be rapidly transported to the ocean via the shallow hydrologic system (*42*). Deep lateral groundwater flow can also transport this otherwise sequestered carbon directly to the ocean via submarine groundwater discharge (SGD) [e.g., (*43*)]; the low salinity estimated for the sediment groundwater directly beneath the ocean cavity at WGZ (Fig. 2D) suggests that SGD may be occurring [Fig. 3C and (*27*)]. Thus, deep groundwater reservoirs beneath ice streams not only contain saline water indicative of past marine incursions but also likely contain the marine microorganisms and carbon that accumulated when the marine sediments were deposited. Groundwater flow can facilitate the mobilization of sequestered carbon to the ice base and oceans during times of enhanced groundwater exfiltration, thereby increasing Antarctica's contribution of carbon to the ocean. The introduction of water into sub–ice shelf cavities via SGD may also affect ocean circulation and dynamics.

For the past several decades, our understanding of Antarctic sub–ice stream water and its relationship to ice behavior has focused on exchange and transport in the shallow hydrologic environment, whereas the presence and role of deeper groundwater have remained largely unexplored because of the lack of observations. Our MT data provide

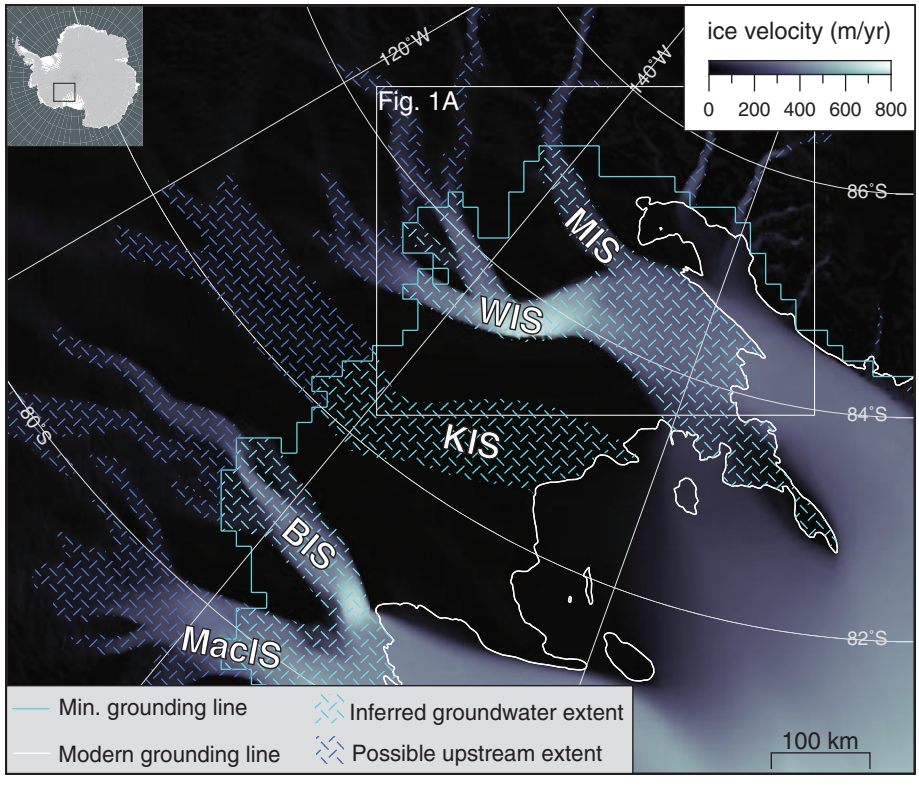


Fig. 4. Potential extent of sub–ice stream deep groundwater systems underlying the Siple Coast. We hypothesize that deep groundwater systems are present within the sedimentary basins underlying Mercer Ice Stream (MIS), WIS, Kamb Ice Stream (KIS), Bindschadler Ice Stream (BIS), and MacAyeal Ice Stream (MacIS). We infer an increase in groundwater salinity with depth (cyan hatching) that possibly extends upstream to the onset of ice streaming (blue hatching).

unequivocal evidence and new constraints for groundwater in a >1-km-thick sedimentary basin. The inferred salinity gradient shows that groundwater represents an active component of the sub–ice stream hydrologic system, redistributing water, heat, solutes, and carbon below WIS. We expect that similar groundwater systems exist within other marine sedimentary basins that underlie Antarctic ice streams, particularly those that have experienced similar episodes of grounding line retreat and readvance, such as in the Siple Coast and Weddell Sea sector (*34*). Understanding the influence of this groundwater on ice sheet behavior will require its integration into the next generation of ice sheet models.

REFERENCES AND NOTES

1. M. R. Bennett, *Earth Sci. Rev.* **61**, 309–339 (2003).
2. M. Bougamont, S. Price, P. Christoffersen, A. J. Payne, *J. Geophys. Res.* **116**, F04018 (2011).
3. D. D. Blankenship, C. R. Bentley, S. T. Rooney, R. B. Alley, *Nature* **322**, 54–57 (1986).
4. K. Christianson, R. W. Jacobel, H. J. Horgan, S. Anandakrishnan, R. B. Alley, *Earth Planet. Sci. Lett.* **331–332**, 237–245 (2012).
5. B. Kamb, in *The West Antarctic Ice Sheet Behavior and Environment*, R. B. Alley, R. A. Bindschadle, Eds. (Antarctic Research Series, vol. 77, American Geophysical Union, 2001), pp. 157–199.
6. M. J. Siegert *et al.*, *Spec. Publ. Geol. Soc. London* **461**, 197–213 (2018).
7. R. E. Bell *et al.*, *Nature* **394**, 58–62 (1998).
8. S. Anandakrishnan, D. D. Blankenship, R. B. Alley, P. L. Stoffa, *Nature* **394**, 62–65 (1998).
9. M. Studinger *et al.*, *Geophys. Res. Lett.* **28**, 3493–3496 (2001).
10. C. Gustafson, H. Fricker, K. Key, M. Siegfried, Wideband magnetotelluric responses from Whillans Ice Stream, West Antarctica. U.S. Antarctic Program (USAP) Data Center (2022); https://doi.org/10.15784/601526.
11. P. Winberry, S. Anandakrishnan, D. Wiens, Geophysical study of ice stream stick-slip dynamics. International Federation of Digital Seismograph Networks (2010); https://doi.org/10.7914/SN/2C_2010.
12. S. Tulaczyk, I. Joughin, S. Schwartz, J. Walter, Elevation change anomalies in West Antarctica and dynamics of subglacial water beneath ice streams and their tributaries. International Federation of Digital Seismograph Networks (2010); https://doi.org/10.7914/SN/1D_2010.
13. T. M. Kyrke-Smith, A. C. Fowler, *Proc. Math. Phys. Eng. Sci.* **470**, 20140340 (2014).
14. R. Kirsch, Ed., *Groundwater Geophysics: A Tool for Hydrogeology* (Springer, 2006).
15. K. Key, M. R. Siegfried, *J. Glaciol.* **63**, 755–771 (2017).
16. N. Foley *et al.*, *Hydrology* **6**, 54 (2019).
17. A. D. Chave, A. G. Jones, Eds., *The Magnetotelluric Method Theory and Practice* (Cambridge Univ. Press, 2012).
18. G. J. Hill, *Surv. Geophys.* **41**, 5–45 (2019).
19. P. Christoffersen, M. Bougamont, S. P. Carter, H. A. Fricker, S. Tulaczyk, *Geophys. Res. Lett.* **41**, 2003–2010 (2014).
20. H. A. Fricker, T. Scambos, *J. Glaciol.* **55**, 303–315 (2009).
21. H. J. Horgan *et al.*, *Geology* **41**, 1159–1162 (2013).
22. H. J. Horgan *et al.*, *Earth Planet. Sci. Lett.* **331–332**, 201–209 (2012).
23. B. C. Christner *et al.*, *Nature* **512**, 310–313 (2014).
24. C. B. Begeman *et al.*, *J. Geophys. Res. Oceans* **123**, 7438–7452 (2018).
25. A. B. Michaud *et al.*, *Geology* **44**, 347–350 (2016).
26. H. J. Horgan, K. Christianson, R. W. Jacobel, S. Anandakrishnan, R. B. Alley, *Geophys. Res. Lett.* **40**, 3934–3939 (2013).
27. Materials and methods are available as supplementary materials.
28. A. Muto, K. Christianson, H. J. Horgan, S. Anandakrishnan, R. B. Alley, *J. Geophys. Res. Solid Earth* **118**, 4535–4546 (2013).
29. R. M. McKay *et al.*, "Expedition 374 preliminary report: Ross Sea West Antarctica Ice Sheet history" (International Ocean Discovery Program, 2018).
30. C. B. Begeman, S. M. Tulaczyk, A. T. Fisher, *Geophys. Res. Lett.* **44**, 9823–9832 (2017).
31. S. Tulaczyk *et al.*, *Ann. Glaciol.* **55**, 51–58 (2014).
32. J. C. Priscu *et al.*, *Ann. Glaciol.* **62**, 340–352 (2021).
33. R. A. Venturelli *et al.*, *Geophys. Res. Lett.* **47**, e2020GL088476 (2020).
34. J. Kingslake *et al.*, *Nature* **558**, 430–434 (2018).
35. V. E. A. Post, H. Kooi, *Hydrogeol. J.* **11**, 549–559 (2003).
36. R. P. Scherer, *Global Planet. Change* **4**, 395–412 (1991).
37. J. M. Lemieux, E. A. Sudicky, W. R. Peltier, L. Tarasov, *J. Geophys. Res.* **113**, F01011 (2008).
38. B. T. Gooch, D. A. Young, D. D. Blankenship, *Geochem. Geophys. Geosyst.* **17**, 395–409 (2016).
39. A. B. Michaud *et al.*, *Nat. Geosci.* **10**, 582–586 (2017).
40. I. P. G. Marshall, S. M. Karst, P. H. Nielsen, B. B. Jørgensen, *Mar. Genomics* **37**, 58–68 (2018).
41. T. J. Vick-Majors *et al.*, *Global Biogeochem. Cycles* **34**, 1–17 (2020).
42. S. P. Carter, H. A. Fricker, *Ann. Glaciol.* **53**, 267–280 (2012).
43. T. Uemura, M. Taniguchi, K. Shibuya, *Geophys. Res. Lett.* **38**, L08402 (2011).
44. J. Mouginot, E. Rignot, B. Scheuchl, *Geophys. Res. Lett.* **46**, 9710–9718 (2019).
45. T. A. Scambos, T. M. Haran, M. A. Fahnestock, T. H. Painter, J. Bohlander, *Remote Sens. Environ.* **111**, 242–257 (2007).
46. M. A. Depoorter *et al.*, *Nature* **502**, 89–92 (2013).

ACKNOWLEDGMENTS

We thank the Antarctic Support Contract, Kenn Borek Air, New York Air National Guard for logistical support; M. Siefert, our field mountaineer, for ensuring the safety of our field camp and assistance with all MT data collection; Phoenix Geophysics for loaning us MT instrument systems and P. Wannamaker for renting us his electrode and buffer preamplifier systems; D. Blatter for assistance with the Bayesian inversion code; A. Muto and K. Christianson for sharing gravity and radar datasets to assist with our interpretation; S. Adusumilli, B. Kulessa, R. Bell, M. Steckler, and the SALSA Science Team for useful discussions; and S. Naif for initial discussions about the prospect of using MT for shallow subglacial science. **Funding:** This work was funded by National Science Foundation grant OPP-1643917 (K.K., C.D.G., M.R.S., and H.A.F.); National Science Foundation grant OPP-0944794 (J.P.W.); National Science Foundation grant OPP-1914767 (J.P.W.); and the Columbia University Electromagnetic Methods Research Consortium (C.D.G.). **Author contributions:** Conceptualization: C.D.G., K.K., M.R.S., J.P.W., H.A.F.; Methodology: C.D.G., K.K., J.P.W.; Software: C.D.G., K.K., J.P.W.; Validation: C.D.G., K.K., J.P.W.; Formal analysis: C.D.G., K.K., J.P.W.; Investigation: C.D.G., K.K., M.R.S., J.P.W.; Resources: K.K., J.P.W.; Data curation: C.D.G., K.K., J.P.W.; Writing – original draft: C.D.G., K.K., M.R.S., J.P.W., H.A.F.; Writing – review and editing: All authors; Visualization: C.D.G., K.K., M.R.S., J.P.W., H.A.F.; Supervision: K.K., M.R.S., J.P.W., H.A.F.; Project administration: C.D.G., K.K., M.R.S., J.P.W.; Funding acquisition: K.K., M.R.S., J.P.W., H.A.F. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All MT data that were inverted and analyzed in this study are available at the US Antarctic Program Data Center (*10*). All seismic data analyzed in this study are archived at the Incorporated Research Institutions for Seismology Data Management Center (IRIS DMC) under network codes 2C (*11*) and 1D (*12*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm3301
Materials and Methods
Figs. S1 to S7
Tables S1 to S3
References (*47–75*)

Submitted 22 September 2021; accepted 2 March 2022
10.1126/science.abm3301

ECOTOXICOLOGY

Conversion of oxybenzone sunscreen to phototoxic glucoside conjugates by sea anemones and corals

Djordje Vuckovic[1], Amanda I. Tinoco[2], Lorraine Ling[2], Christian Renicke[2], John R. Pringle[2], William A. Mitch[1]*

The reported toxicity of oxybenzone-based sunscreens to corals has raised concerns about the impacts of ecotourist-shed sunscreens on corals already weakened by global stressors. However, oxybenzone's toxicity mechanism(s) are not understood, hampering development of safer sunscreens. We found that oxybenzone caused high mortality of a sea anemone under simulated sunlight including ultraviolet (UV) radiation (290 to 370 nanometers). Although oxybenzone itself protected against UV-induced photo-oxidation, both the anemone and a mushroom coral formed oxybenzone–glucoside conjugates that were strong photo-oxidants. Algal symbionts sequestered these conjugates, and mortality correlated with conjugate concentrations in animal cytoplasm. Higher mortality in anemones that lacked symbionts suggests an enhanced risk from oxybenzone to corals bleached by rising temperatures. Because many commercial sunscreens contain structurally related chemicals, understanding metabolite phototoxicity should facilitate the development of coral-safe products.

Most of the world's coral reefs are endangered (*1*, *2*). Although much of the threat is due to global factors, including increasing sea temperatures (*3*), coral declines can be exacerbated by local, anthropogenic factors. Much recent concern has focused on oxybenzone, a common sunscreen ingredient. For example, research in the US Virgin Islands found no substantial settlement of coral larvae, survival of juvenile corals, or regeneration of adult tissue in induced lesions over a 5-year period in Trunk Bay, where high levels of recreational swimming resulted in up to 1.4 mg of oxybenzone per liter of seawater (*4*). Meanwhile, a thriving

[1]Department of Civil and Environmental Engineering, Stanford University, Stanford, CA, USA. [2]Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA.
***Corresponding author. Email: wamitch@stanford.edu**

coral community was found at neighboring Caneel Bay, with lower recreational use but presumably the same impacts from global stressors. Exacerbation of coral declines by sunscreens washed off tourists would be ironic and particularly pernicious, given the promotion of ecotourism in the interest of protecting coral reefs.

Although previous research has indicated that oxybenzone is an endocrine disruptor (*5*, *6*), studies using corals and zebrafish have suggested that it might also be phototoxic (*4*, *7*). However, the mechanism(s) of phototoxicity have not been explained. Here, we demonstrate that oxybenzone is indeed a sunscreen in vitro but is metabolized into phototoxic glucoside conjugates in the sea anemone *Aiptasia* and the mushroom coral *Discosoma* sp. (see materials and methods and fig. S1). We also show that *Aiptasia* without algal symbionts accumulate more glucoside metabolites in their tissues and die faster than animals with algae, suggesting that the algae protect the hosts by sequestering the phototoxins. Thus, oxybenzone may be particularly phototoxic to bleached corals, exacerbating the damage due to global stressors. With recent moves by regulatory authorities in Hawaii and elsewhere to ban oxybenzone, understanding the mechanism(s) of its phototoxicity is important to ensure that the sunscreen components that are selected as alternatives are truly safer for corals.

For most experiments, we used *Aiptasia*, a well-established model system for the study of symbiotic anthozoan cnidarians such as corals (see materials and methods and table S1). We exposed symbiotic anemones to 8.8 μM (2 mg/liter) oxybenzone in artificial seawater (ASW) at 27°C in a solar simulator that approximates the 24-hour diurnal sunlight cycle (see materials and methods and fig. S2), including ultraviolet (UV) wavelengths of 290 to 370 nm (Fig. 1A). Mortality was 100% within 17 days (Fig. 1B and table S2). By contrast, negligible mortality was observed over 21 days when the animals were exposed to simulated sunlight without oxybenzone or to 8.8 μM oxybenzone without UV light (Fig. 1B).

The importance of 290- to 370-nm light for oxybenzone-induced phototoxicity was surprising, given that oxybenzone's strong absorption within this waveband (Fig. 2, A and B) is precisely why it is expected to protect sunscreen users from UV damage. Phototoxicity frequently involves the photoexcitation of sensitizing molecules to excited triplet states that then degrade biomolecules either directly or indirectly via reactive oxygen species (ROS) and/or reactive halogen species (RHS) (*8*). To test this possibility, we conducted in vitro experiments in a solution that contained seawater levels of halides using allyl-thiourea and sorbic alcohol as probes for photosensitization by oxybenzone (Fig. 2C), because their reduced sulfur and alkene groups are reactive with excited-state triplets, ROS, and RHS (*8*, *9*). Consistent with oxybenzone's intended role as a sunscreen and with previous research indicating that it does not produce ROS upon sunlight illumination in vitro (*10*), oxybenzone decreased the first-order photodegradation rate of allyl-thiourea by ~40% (Fig. 2D, none versus **1**; P = 0.006; fig. S3; and table S3); similar results were observed with sorbic alcohol (fig. S4).

Because oxybenzone is similar in structure to the strong photosensitizer benzophenone (BP) (Fig. 2A, **1** and **6**) (*11*), we hypothesized that *Aiptasia* (and also corals) might metabolize oxybenzone to one or more photosensitizers. Previous studies of oxybenzone metabolites in corals only analyzed for 2,4-dihydroxybenzophenone and 2,2′-dihydroxy-4-methoxybenzophenone and either observed them at very low (~0.04 nmol/g wet weight) concentrations (*12*) or did not observe them at all (*13*). However, our analysis of *Aiptasia* tissue (see materials and methods) after exposure to 8.8 μM oxybenzone for 18 hours found that these compounds accounted for <2% of total metabolite mass. We instead recovered 1780 nmol/g dry weight (~220 nmol/g wet weight) of oxybenzone-glucoside (Fig. 2A, **4**) and higher-order conjugates of oxybenzone (Fig. 3, A and B; figs. S5 to S7; and table S3).

Glucoside conjugates of oxybenzone were reported previously only in plants and suggested to be a pathway for oxybenzone detoxification (*14*). Here, however, whereas oxybenzone protected allyl-thiourea from photodegradation in vitro (Fig. 2D), its glucoside metabolite increased the photodegradation rate by ~5.6-fold (Fig. 2D, none versus **4**; P < 0.001). This photodegradation occurred only during illumination with the 290- to 370-nm waveband (fig. S3B), which corresponds to the glucoside's absorption spectrum (Fig. 2B). The glucoside's concentration did not change during the reaction (fig. S3C), indicating that it generates reactive species photocatalytically. We also observed enhanced allyl-thiourea photodegradation when the hydroxyl group that is adjacent to the carbonyl was deprotonated at pH 11 or replaced by a methoxy group [$(MeO)_2$BP] [**2** (P < 0.0001) and **3** (P < 0.001) in Fig. 2, A and D]; similar results were observed when sorbic alcohol was used as a probe (fig. S4). These results agree with previous ab initio calculations (*15*) and pump-probe spectroscopy (*16*) results indicating the importance of the hydroxyl group for sunscreen activity; transfer of the proton from the hydroxyl group to the photoexcited triplet carbonyl promotes its relaxation to the ground state, which should prevent damage to biomolecules.

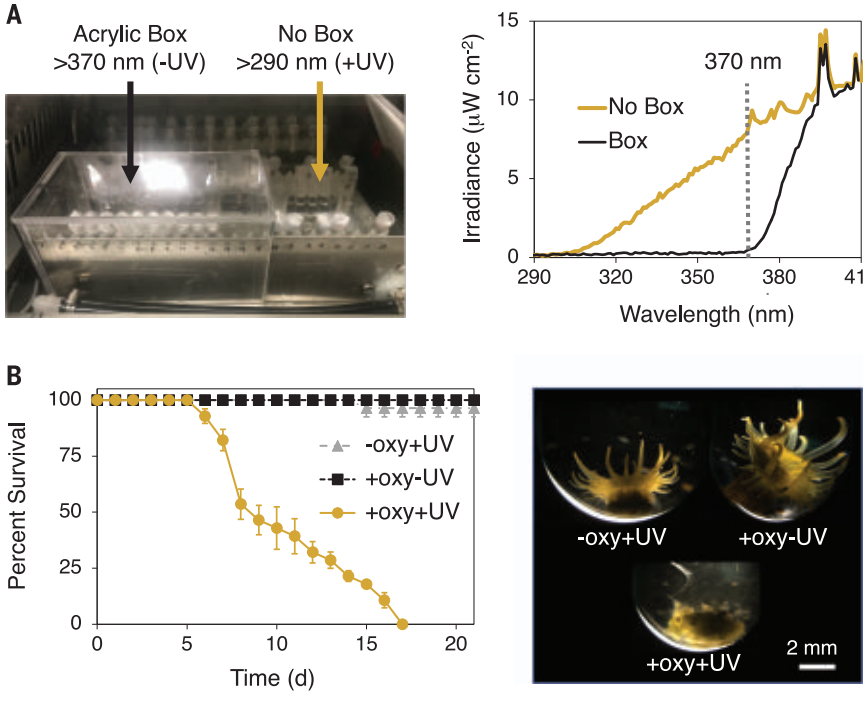


Fig. 1. Phototoxicity of oxybenzone to *Aiptasia* under simulated sunlight. (**A**) Solar simulator setup. An acrylic box (left) blocks UV light of wavelength <370 nm (right). (**B**) Survival of symbiotic *Aiptasia* in ASW exposed to full-spectrum light without oxybenzone (−oxy+UV; n = 28 animals in total) or to 8.8 μM oxybenzone either with full-spectrum light (+oxy+UV; n = 28 animals) or with light >370 nm (+oxy−UV; n = 16 animals). Animals were exposed in three independent trials; means ± SEM (weighted for the number of animals in each trial) are shown. Photographs show an example from each group after 9 days.

The high concentration of oxybenzone-glucoside and other presumably photosensitizing metabolites in *Aiptasia* co-occurred with oxybenzone itself (Fig. 3B, whole animal, oxy). It was not possible to determine the phototoxicities of the metabolites directly by introducing them into *Aiptasia* through the water column because of their low lipophilicity (octanol–water partition coefficient log K_{ow} = 1.1 for oxybenzone-glucoside versus 3.6 for oxybenzone). Thus, we treated *Aiptasia* with $(MeO)_2BP$, which provides photosensitization comparable to that of oxybenzone-glucoside (Fig. 2D, **3** versus **4**; P = 0.35) but readily partitions into *Aiptasia* from the water column (log K_{ow} = 3.3). Exposure to 8.3 μM $(MeO)_2BP$ caused little to no mortality in the absence of 290- to 370-nm illumination [Fig. 3C; compare with the absorption spectrum of $(MeO)_2BP$, Fig. 2B]. However, under full-spectrum light, 8.3 μM $(MeO)_2$-BP caused rapid mortality, and even 0.83 μM $(MeO)_2$-BP caused mortality at a rate threefold faster than 8.8 μM oxybenzone (Fig. 3C; $P = 1 \times 10^{-10}$).

Despite the greater lethality of $(MeO)_2BP$, *Aiptasia* exposed to this molecule at 0.83 μM contained ~22-fold less phototoxin than *Aiptasia* exposed to 8.8 μM oxybenzone [Fig. 3B, whole animal, $(MeO)_2$-BP versus sum of oxybenzone-glucoside and higher–molecular weight (MW) conjugates ($P = 6 \times 10^{-9}$)]. The fractionation of *Aiptasia* homogenates (see materials and methods) suggested two explanations for this apparent paradox. First, the glucoside metabolites of oxybenzone were sequestered either entirely (higher–molecular weight conjugates, such as the oxy-malonyl-glucoside) or largely (the glucoside without further modification) within the algal symbionts (Fig. 3B, Algal versus Animal fraction), so that the animal tissue of *Aiptasia* exposed to oxybenzone contained only ~1.5-fold more phototoxin than the animal tissue of *Aiptasia* exposed to $(MeO)_2BP$ [Fig. 3B, inset: oxy-glucoside versus $(MeO)_2BP$; P = 0.03]. Second, the animal tissue of oxybenzone-treated *Aiptasia* also contained a substantial amount of oxybenzone itself (Fig. 3B, Animal fraction), which probably provided a UV-screening effect that was absent from *Aiptasia* treated with $(MeO)_2BP$.

These results suggested that the algae protect the animals by sequestering phototoxic oxybenzone metabolites and that the overall phototoxicity depends on the balance in the animal tissue between screening of UV light by oxybenzone and photosensitization by its glucoside metabolites. Support for this model was obtained from experiments with aposymbiotic (lacking algae) *Aiptasia* (see materials and methods) and with the mushroom coral *Discosoma*. First, aposymbiotic *Aiptasia* died much faster in the presence of 8.8 μM oxybenzone than symbiotic *Aiptasia* (Fig. 4A; $P = 6 \times 10^{-12}$), and correspondingly, the aposymbiotic tissue contained ~2.7-fold more oxybenzone-glucoside than the animal fraction of symbiotic anemones (Fig. 4C; $P = 1 \times 10^{-6}$). Second, *Discosoma* showed no death during 8 days under full-spectrum light with 8.8 μM oxybenzone (fig. S8), even though they were as susceptible as *Aiptasia* to the phototoxin $(MeO)_2BP$ (Fig. 4B; P = 0.2). Correspondingly, tissue analyses indicated that the algae in *Discosoma* were even more effective than those in *Aiptasia* at sequestering the oxybenzone metabolites, which resulted in undetectable concentrations of the phototoxins within the surrounding animal tissue (Fig. 4C). Relative to *Aiptasia*, the *Discosoma* samples contained lower numbers of algal cells per unit host protein (fig. S1C; $P = 6 \times 10^{-6}$). Thus, the difference in phototoxin sequestration probably reflects distinct intrinsic properties of the different algal species in the two hosts (fig. S1), so that the exchange of algal symbionts that may occur during stress and subsequent recovery (*17*) could alter susceptibility to oxybenzone-induced phototoxicity.

If the symbiotic algae of corals and other anthozoans indeed protect them from the toxic effects of oxybenzone metabolites, then the widespread bleaching of corals in response to rising seawater temperatures (*2*) will make

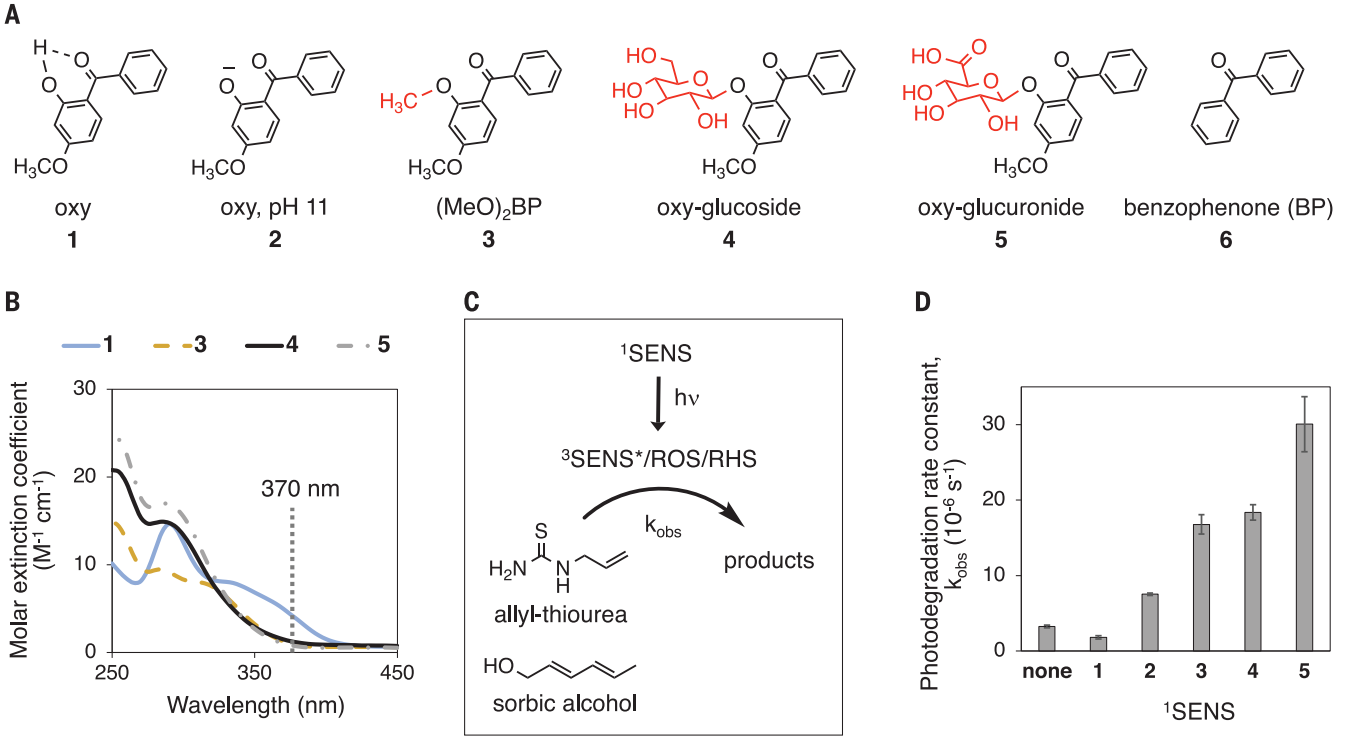


Fig. 2. Photosensitization by oxybenzone and related molecules. (**A**) Chemical structures of oxybenzone (oxy) and related molecules (red: structural additions to oxybenzone). (**B**) Light-absorption spectra of compounds **1** and **3** to **5**. (**C**) Allyl-thiourea and sorbic alcohol as probes for reactive species. ^{1}SENS, photosensitizing molecule in its ground state; ^{3}SENS*, photosensitizing molecule in its excited, triplet state. (**D**) First-order photodegradation rate constants for 10 μM allyl-thiourea with 35 μM of compounds **1** to **5** at pH 8.1 in seawater-strength halide solution (see supplementary materials and methods). Error bars, SEM from three independent experiments.

them more susceptible to oxybenzone-induced phototoxicity. This may be a practical problem at reefs where recreational swimming and sunscreen use are high, with the risk varying between species. The widespread occurrence of glucosides and related metabolites suggests that phototoxicity from oxybenzone might also affect species beyond anthozoans. Notably, oxybenzone-glucuronide, the dominant oxybenzone metabolite in humans (*18*), is as potent a photosensitizer as oxybenzone-glucoside (Fig. 2D and fig. S4). Thus, metabolic pathways that presumably evolved to detoxify contaminants by rendering them more water soluble apparently convert oxybenzone from a sunscreen to

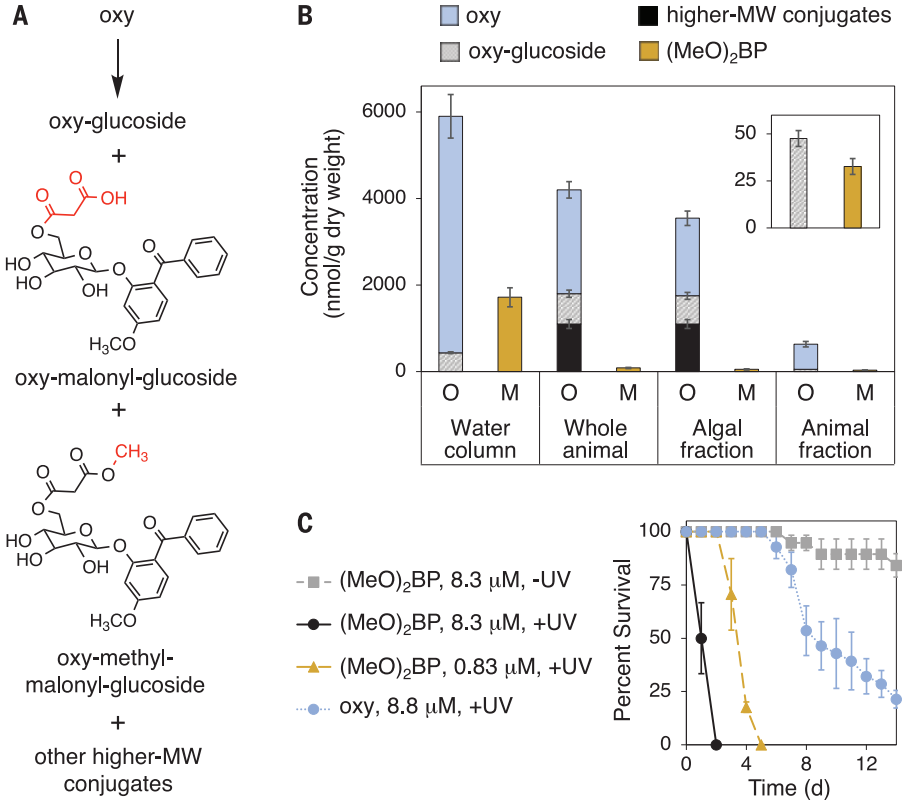


Fig. 3. Oxybenzone metabolism in *Aiptasia* and $(MeO)_2BP$ effects. (**A**) Hypothesized scheme of oxybenzone (oxy) metabolism. (**B**) Concentrations of oxybenzone, its glucoside metabolites, and $(MeO)_2BP$ (Fig. 2A, **3**) in the water column, whole *Aiptasia*, and its algal and animal fractions after an 18-hour exposure to 8.8 μM oxybenzone (O) or 0.83 μM $(MeO)_2BP$ (M). (Inset) Concentrations of oxybenzone-glucoside and $(MeO)_2BP$ in the animal fraction (note expanded scale). Error bars: SEM of six trials (two animals per trial). (**C**) Survival of *Aiptasia* exposed to the full solar spectrum (+UV) with 8.3 μM $(MeO)_2BP$ (n = 16), 0.83 μM $(MeO)_2BP$ (n = 17), or 8.8 μM oxybenzone (data from Fig. 1B) or to light >370 nm (−UV) with 8.3 μM $(MeO)_2BP$ (n = 19). Error bars, SEM over three independent trials, weighted for the number of anemones in each trial.

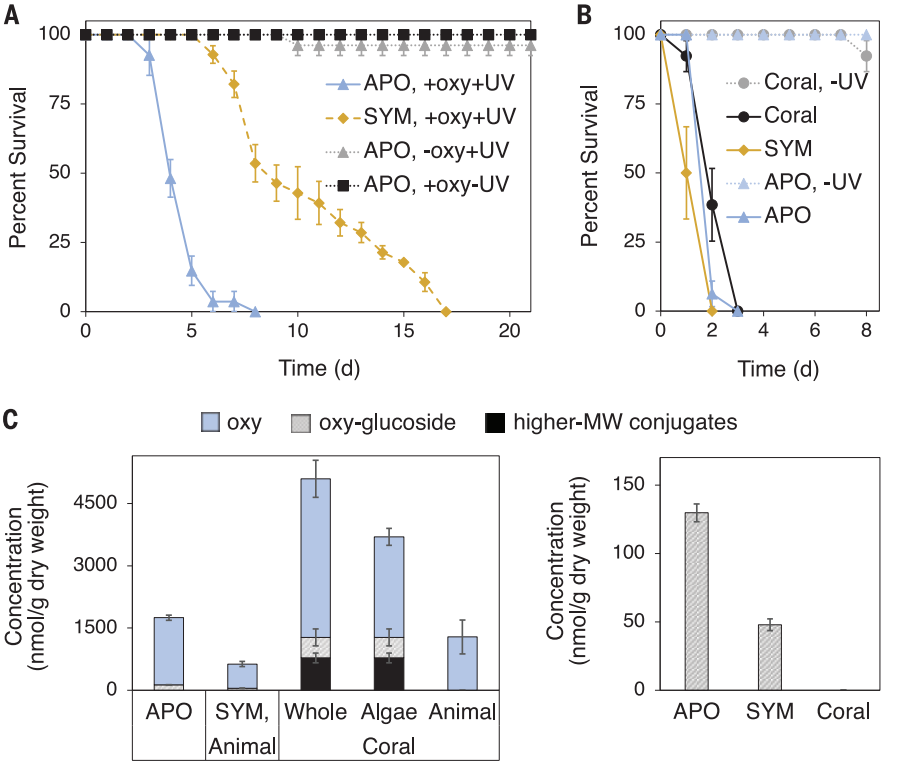


Fig. 4. Oxybenzone sensitivity and metabolite concentrations in aposymbiotic *Aiptasia* and in the mushroom coral *Discosoma*. (**A**) Survival of aposymbiotic *Aiptasia* (APO) in the presence of 8.8 μM oxybenzone and full-spectrum light (+oxy+UV; n = 27), full-spectrum light without oxybenzone (−oxy+UV; n = 26), or 8.8 μM oxybenzone without light <370 nm (+oxy−UV; n = 16). Survival data for symbiotic *Aiptasia* (SYM) from Fig. 1B are included for comparison. Error bars, SEM over three trials (weighted by anemones per trial). (**B**) Survival of *Discosoma* coral (n = 13), SYM *Aiptasia* (n = 16, data from Fig. 3C), and APO *Aiptasia* (n = 16) exposed to 8.3 μM $(MeO)_2BP$ under full-spectrum light and of APO *Aiptasia* (n = 8) and *Discosoma* (n = 13) exposed to 8.3 μM $(MeO)_2BP$ without light <370 nm (−UV). (**C**) Left: oxybenzone and metabolite concentrations in the animal fractions of SYM (from Fig. 3B) and APO *Aiptasia* and in the whole organism and algal and animal fractions of *Discosoma* after exposure to 8.8 μM oxybenzone for 18 hours. Right: concentrations of oxybenzone-glucoside in the animal fractions replotted on an expanded scale. Error bars, SEM over six trials for SYM and APO and three trials for *Discosoma* (two animals per trial).

potent phototoxins. These same pathways may also produce phototoxins from other organic sunscreens [e.g., salicylates (*19*) and other benzophenones (*20*)], which have aromatic carbonyl structures similar to that in oxybenzone and also rely on an excited-state proton transfer as their main energy-dissipation mechanism. Such conversions will need to be taken into account in developing safer alternative sunscreens.

REFERENCES AND NOTES

1. T. P. Hughes *et al*., *Nature* **546**, 82–90 (2017).
2. T. P. Hughes *et al*., *Science* **359**, 80–83 (2018).
3. S. F. Heron *et al*., "Impacts of climate change on world heritage coral reefs: a first global scientific assessment" (UNESCO World Heritage Centre, 2017); https://whc.unesco.org/document/158688.
4. C. A. Downs *et al*., *Arch. Environ. Contam. Toxicol.* **70**, 265–288 (2016).
5. S. Kim, K. Choi, *Environ. Int.* **70**, 143–157 (2014).
6. M. Ghazipura, R. McGowan, A. Arslan, T. Hossain, *Reprod. Toxicol.* **73**, 175–183 (2017).
7. Y. Zhang *et al*., *Aquat. Toxicol.* **235**, 105835 (2021).
8. K. M. Parker, W. A. Mitch, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 5868–5873 (2016).
9. H. Zhou, S. Yan, J. Ma, L. Lian, W. Song, *Environ. Sci. Technol.* **51**, 11066–11074 (2017).
10. J. M. Allen, C. J. Gossett, S. K. Allen, *Chem. Res. Toxicol.* **9**, 605–609 (1996).
11. M. C. Cuquerella, V. Lhiaubet-Vallet, J. Cadet, M. A. Miranda, *Acc. Chem. Res.* **45**, 1558–1570 (2012).
12. M. M. P. Tsui *et al*., *Environ. Sci. Technol.* **51**, 4182–4190 (2017).
13. C. L. Mitchelmore *et al*., *Sci. Total Environ.* **670**, 398–410 (2019).
14. F. Chen, C. Huber, R. May, P. Schröder, *J. Hazard. Mater.* **306**, 230–236 (2016).
15. T. N. V. Karsili, B. Marchetti, M. N. R. Ashfold, W. Domcke, *J. Phys. Chem. A* **118**, 11999–12010 (2014).
16. L. A. Baker *et al*., *J. Phys. Chem. Lett.* **6**, 1363–1368 (2015).
17. R. Cunning, R. N. Silverstein, A. C. Baker, *Coral Reefs* **37**, 145–152 (2018).
18. X. Ye, Z. Kuklenyik, L. L. Needham, A. M. Calafat, *Anal. Bioanal. Chem.* **383**, 638–644 (2005).
19. E. L. Holt *et al*., *Phys. Chem. Chem. Phys.* **22**, 15509–15519 (2020).
20. L. A. Baker, B. Marchetti, T. N. V. Karsili, V. G. Stavros, M. N. R. Ashfold, *Chem. Soc. Rev.* **46**, 3770–3791 (2017).

ACKNOWLEDGMENTS

We thank M. Houlihan (survival experiments), A. Simpson, E. Sattely, and K. Smith (metabolite purification), M. Paul (protein analysis), G. Rosenfield (coral harvesting and advice), O. Barry (animal feeding), B. Sabala (test-tube rack), T. Xiang (algal cultures), J. Bolorinos (statistics advice), and A. Grossman and Mitch and Pringle laboratory members for thoughtful insights throughout the project. **Funding:** Woods Institute for the Environment at Stanford University (grant no. WTAWZ to J.R.P. and W.A.M.); National Science Foundation (NSF) Graduate Research Fellowship (to D.V.); and NSF (grant no. CBET-2114790 to W.A.M. and J.R.P.). **Author contributions:** Conceptualization: J.R.P., W.A.M., and D.V.; Methodology: D.V., A.I.T., L.L., C.R., J.R.P., and W.A.M.; Investigation: D.V., A.I.T., L.L., and C.R.; Funding acquisition: D.V., J.R.P., and W.A.M.; Supervision: J.R.P. and W.A.M.; Writing—original draft: D.V. and W.A.M.; Writing—review and editing: D.V., A.I.T., L.L., C.R., J.R.P., and W.A.M. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** All data are available in the main text or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abn2600
Materials and Methods
Figs. S1 to S8
Tables S1 to S4
References (*21*–*36*)
MDAR Reproducibility Checklist
Data S1 to S10

Submitted 15 November 2021; accepted 16 March 2022
10.1126/science.abn2600

FERROELECTRICS

Emergent ferroelectricity in subnanometer binary oxide films on silicon

Suraj S. Cheema[1,2]*†, Nirmaan Shanker[2]†, Shang-Lin Hsu[2]†, Yoonsoo Rho[3], Cheng-Hsiang Hsu[2], Vladimir A. Stoica[4,5], Zhan Zhang[5], John W. Freeland[5], Padraic Shafer[6], Costas P. Grigoropoulos[3], Jim Ciston[7], Sayeef Salahuddin[2,8]*

The critical size limit of voltage-switchable electric dipoles has extensive implications for energy-efficient electronics, underlying the importance of ferroelectric order stabilized at reduced dimensionality. We report on the thickness-dependent antiferroelectric-to-ferroelectric phase transition in zirconium dioxide (ZrO_2) thin films on silicon. The emergent ferroelectricity and hysteretic polarization switching in ultrathin ZrO_2, conventionally a paraelectric material, notably persists down to a film thickness of 5 angstroms, the fluorite-structure unit-cell size. This approach to exploit three-dimensional centrosymmetric materials deposited down to the two-dimensional thickness limit, particularly within this model fluorite-structure system that possesses unconventional ferroelectric size effects, offers substantial promise for electronics, demonstrated by proof-of-principle atomic-scale nonvolatile ferroelectric memory on silicon. Additionally, it is also indicative of hidden electronic phenomena that are achievable across a wide class of simple binary materials.

The evolution of ferroic order at reduced dimensions—in particular, two-dimensional (2D) ferroelectricity—has long been intriguing for scaled energy-efficient electronics (*1*) because of the inherent ability to control electric polarization with an applied voltage (*2*). Since the discovery of ferroelectricity and antiferroelectricity in HfO_2-ZrO_2–based thin films (*3*, *4*), fluorite-structure binary oxides have reignited such interest (*5*) because they overcome many of the thickness-scaling (*6*, *7*) and silicon-compatibility issues (*8*) that afflict their perovskite and van der Waals ferroelectric counterparts. Early studies into fluorite-structure systems examined the ferroelectric-antiferroelectric phase competition in HfO_2-ZrO_2 solid solutions as a function of composition (*9*), typically in the 10-nm regime (*8*); meanwhile, recent works demonstrated ferroelectricity down to a sub-2-nm thickness in epitaxial (*10*) and polycrystalline (*11*) Zr:HfO_2 films.

Considering the implications of voltage-driven polarization switching for memory applications (*8*), the fundamental size limit of ferroelectric order in fluorite-structure systems is of critical importance. First-principles calculations have shown that 2D HfO_2 layers in their polar orthorhombic structure ($Pca2_1$) have minimal electrostatic penalty—that is, depolarizing field—enabling unsuppressed polarization down to the unit-cell limit (*12*, *13*). Furthermore, monolayer ZrO_2 was predicted to support switchable polarization on an atomically abrupt structure with Si (*14*). These predictions of scale-free fluorite-structure ferroelectricity (*12*, *14*) strongly motivate experimental demonstration of subnanometer polarization switching in this binary oxide system on silicon toward realizing highly scaled low-power nonvolatile memories (*8*, *12*).

Our strategy to achieve atomic-scale ferroelectricity aims to convert the conventionally antiferroelectric tetragonal phase (t-phase) of ZrO_2 (t-ZrO_2: $P4_2/nmc$) to the ferroelectric orthorhombic phase (o-phase) of ZrO_2 (o-ZrO_2: $Pca2_1$) through reduced dimensionality (Fig. 1A). The reduced dimensionality stabilizes the pressure-induced ferroelectric o-phase in fluorite-based oxides—conventionally achieved through hydrostatic pressure (*15*), chemical pressure (*9*), or epitaxial strain (*16*)—in the ultrathin regime, akin to size-driven antiferroelectric-to-ferroelectric phase transitions observed in prototypical perovskite ferroelectrics (*17*). We demonstrate the emergence of atomic-scale ferroelectricity in conventionally paraelectric ZrO_2 films down to a thickness of 5 Å, corresponding to the fluorite-structure unit-cell size.

ZrO_2 films with thicknesses from 10 nm to 5 Å (Fig. 1B), which we confirmed with synchrotron x-ray and transmission electron microscopy (TEM) analysis (figs. S1 and S2), are grown by atomic layer deposition on SiO_2-buffered Si (Fig. 1C) (*18*). To study the thickness-dependent antiferroelectric-ferroelectric evolution in ZrO_2, we examined the structural signatures of the respective t- and o-phases (Fig. 1). Synchrotron in-plane grazing incidence diffraction (IP-GID) spectra confirm the expected t-phase (101)

[1]Department of Materials Science and Engineering, University of California, Berkeley, CA, USA. [2]Department of Electrical Engineering and Computer Sciences, University of California, Berkeley, CA, USA. [3]Laser Thermal Laboratory, Department of Mechanical Engineering, University of California, Berkeley, CA, USA. [4]Department of Materials Science and Engineering, Pennsylvania State University, University Park, PA, USA. [5]Advanced Photon Source, Argonne National Laboratory, Lemont, IL, USA. [6]Advanced Light Source, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [7]National Center for Electron Microscopy, Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [8]Materials Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, CA, USA.
***Corresponding author. Email: s.cheema@berkeley.edu (S.S.C.); sayeef@berkeley.edu (S.S.)**
†These authors contributed equally to this work.

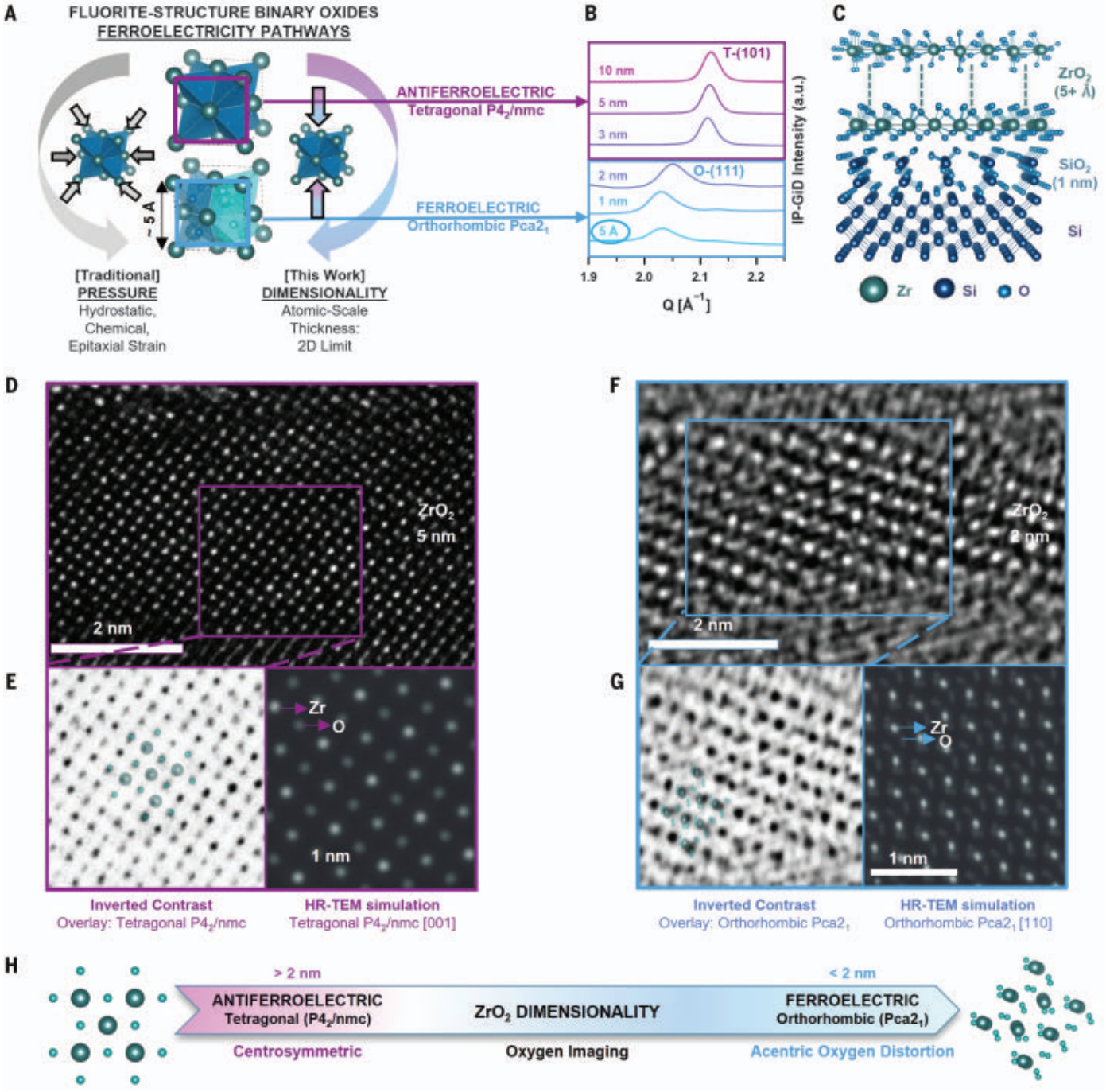


Fig. 1. Atomic-scale emergence of ferroelectricity in ZrO_2. (**A**) Pressure-driven pathways to ferroelectricity in fluorite-structure binary oxides; dimensionality—that is, atomic-scale thickness scaling—can induce ferroelectric o-ZrO_2 stabilization in conventionally antiferroelectric t-ZrO_2. (**B**) Thickness-dependent IP-GID demonstrating the tetragonal-to-orthorhombic phase transition in the ultrathin (5 Å to 2 nm) regime (fig. S4). a.u., arbitrary units; Q, scattering vector. (**C**) Schematic of the Si/SiO_2(1 nm)/ZrO_2(0.5 to 10 nm) films grown by atomic layer deposition. (**D** to **G**) Cross-sectional HR-TEM images of 5-nm ZrO_2 (D) and 2-nm ZrO_2 (F) indexed by oxygen-sensitive negative spherical aberration imaging and inverted contrast HR-TEM simulations (figs. S5 and S6), fit to the t-phase (E) and o-phase (G) lattice along the [001] and [110] zone axes, respectively. For inverted contrast images [(E) and (G)], light (dark) atoms represent O (Zr) atoms. Note the faint zigzag oxygen arrangements expected for the o-phase [110] projection (G). (**H**) Dimensionality-driven antiferroelectric-to-ferroelectric evolution of ZrO_2 demonstrated through oxygen imaging, sensitive to the anion distortion present in fluorite-structure binary oxides.

reflection in thicker (3 to 10 nm) ZrO_2 films and the emergence of the o-phase (111) reflections for ultrathin (≤2 nm) films down to a thickness of 5 Å (Fig. 1B), concurrent with the emergence of highly oriented films (fig. S3) and supported by detailed IP-GID indexing (fig. S4) (*18*).

Distinguishing the nearly identical polar o-phase and nonpolar t-phase structural polymorphs can be further clarified by local oxygen atomic imaging (*19*, *20*). We used oxygen-sensitive negative spherical aberration imaging (*20*) and high-resolution TEM (HR-TEM) to identify the expected nonpolar t-phase in 5-nm ZrO_2 films (Fig. 1, D and E), aided by oxygen

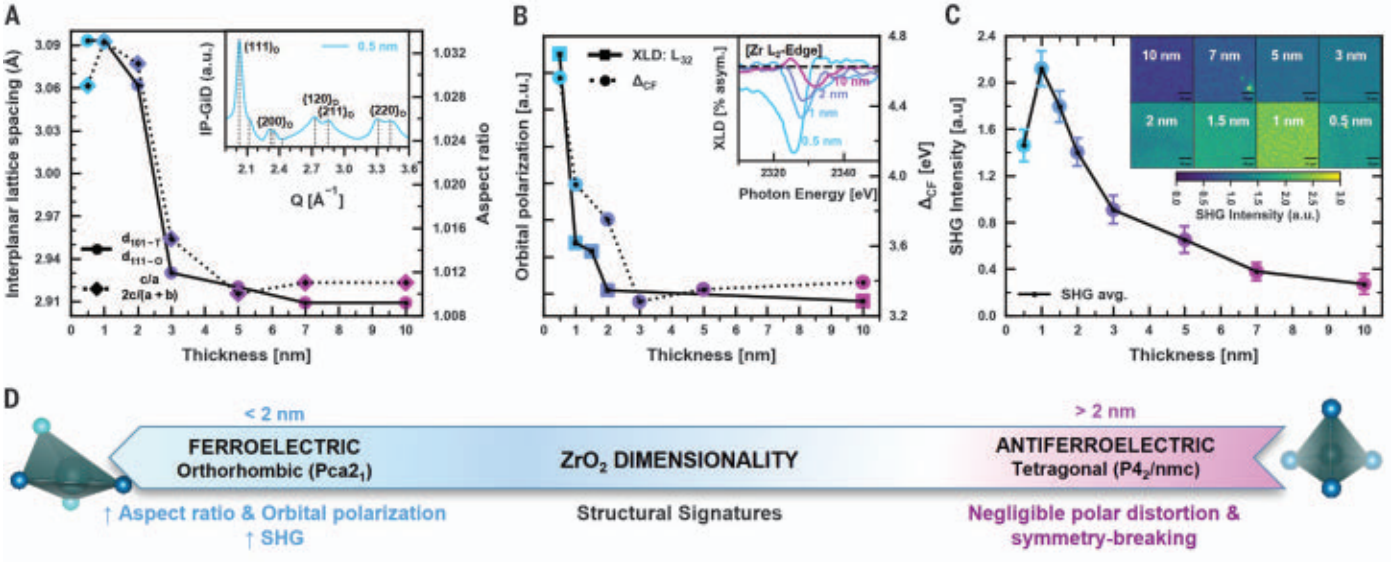


Fig. 2. Thickness-dependent ferroic phase evolution in ultrathin ZrO_2. (**A**) Thickness-dependent lattice spacing (t-d_{101} or o-d_{111}; solid line; left *y* axis) and aspect ratio (t-phase c/a or o-phase $2c/(a + b)$; dashed line; right *y* axis) indicating ultrathin-enhanced lattice distortion. The inset shows example IP-GID spectra for 0.5-nm ZrO_2 indexed to the ferroelectric o-phase. The structural markers for ultrathin (0.5 to 2 nm) and thicker (3 to 10 nm) ZrO_2 films are extracted from IP-GID spectra (fig. S4) and OOP-GID spectra, respectively (*18*). (**B**) Thickness-dependent crystal-field splitting (O *K*-edge x-ray absorption spectra; solid line; left *y* axis) and orbital polarization (Zr-$L_{3,2}$-edge x-ray linear dichroism (XLD); dashed line; right *y* axis) indicating ultrathin-amplified structural and polar distortion (*18*) (fig. S8). The inset shows an increase in orbital polarization at the Zr-L_2-edge with decreasing ZrO_2 thickness. Δ_{CF}, crystal field distortion. (**C**) Averaged SHG intensity of bare ZrO_2 films (0.5 to 10 nm) increases with decreasing ZrO_2 thickness, indicative of ultrathin-enhanced remnant polarization. The inset shows 2D SHG maps for the entire ZrO_2 thickness series. Further evidence of strong (weak) SHG intensity in ultrathin polar (thick nonpolar) ZrO_2 samples is provided by SHG spectra (fig. S11). (**D**) Dimensionality-driven antiferroelectric-to-ferroelectric evolution of ZrO_2 demonstrated through various structural signatures.

analysis (fig. S5) and HR-TEM simulations (fig. S6). Notably, this technique indicates that the polar o-phase emerges in 2-nm ZrO_2 films (Fig. 1, F and G), which x-ray analysis (Fig. 1B) pinpoints as the onset of ferroelectric o-ZrO_2 stabilization, as identified by its characteristic zigzag-like oxygen arrangement that is visible along the [110] projection (Fig. 1G). Additionally, lattice-angle analysis from traditional cation imaging matches $P4_2/nmc$ and $Pca2_1$ HR-TEM simulations for 5- and 2-nm ZrO_2 films, respectively (fig. S7), consistent with oxygen imaging.

To further examine this thickness-dependent antiferroelectric-to-ferroelectric transition, we used additional synchrotron x-ray studies to detect structural signatures of the respective ferroic phases (Fig. 2). From thickness-dependent GID, we found that the interplanar lattice spacing (o-d_{111} or t-d_{101}) and aspect ratio [$2c/(a + b)$ or c/a], which are structural barometers of lattice distortion established for fluorite-structure films (*18*), sharply rise at thicknesses less than 3 nm (Fig. 2A). These observations indicate increased polar o-ZrO_2 stabilization in the atomic-scale limit (Fig. 1B). Along with diffraction, x-ray spectroscopy provides another gauge of the fluorite-structure symmetry. Thickness-dependent x-ray absorption spectra at the oxygen *K*-edge (fig. S8) demonstrate larger crystal field splitting (Fig. 2B) below 3 nm, suggesting more pronounced polar distortion (*18*). Indeed, secondary crystal field splitting that arises from the polar rhombic distortion, a fingerprint of the ferroelectric o-phase (fig. S8), emerges for 5- and 10-Å ZrO_2. Furthermore, x-ray linear dichroism from zirconium $M_{3,2}$- and $L_{3,2}$-edges (fig. S8) indicates more pronounced orbital polarization, linked to electric polarization that arises from ferroelectric polar distortions (*11*, *18*), in the ultrathin regime (Fig. 2B). These thickness-dependent diffraction and spectroscopy trends support the ultrathin-amplified emergence of ferroelectric o-ZrO_2 in typically antiferroelectric t-ZrO_2 films.

Optical microscopy observations also support the size-dependent ferroic phase evolution, because the increase in the second-harmonic generation (SHG) signal, which is related to macroscopic polarization (*18*), with decreasing ZrO_2 thickness (Fig. 2C) is consistent with the ultrathin-enhanced polar distortion trends. Additionally, thickness-dependent capacitance-voltage (*C-V*) measurements of metal-oxide-semiconductor capacitors (fig. S9) indicate a crossover from an antiferroelectric-like t-phase permittivity (κ), $\kappa \sim 40$ (*4*), toward a more ferroelectric-like o-phase permittivity, $\kappa \sim 30$ (*4*, *21*), for ultrathin ZrO_2 films, again consistent with structural characterization.

To further characterize the electrical behavior, we fabricated metal-insulator-metal (MIM) capacitors with varying ZrO_2 thicknesses (Fig. 3A). Considering that antiferroelectrics are defined based on their field-induced transition to a proximal polar phase and not simply their parent crystal structure (*17*), voltage-dependent hysteretic behavior is required to probe the underlying ferroic order, beyond crystallographic signatures of their parent structure. MIM polarization-voltage (*P-V*) loops for 5- and 10-nm-thick ZrO_2, the typical t-ZrO_2 thickness regime (*4*, *8*), demonstrate a signature antiferroelectric-like double hysteresis (*17*) (Fig. 3B). Importantly, conventional *P-V* probes of the signature behavior cannot be applied to the ultrathin regime (*18*), in which nonpolarization-dependent leakage current masks polarization-dependent switching current.

To directly probe the polarization switching properties of ultrathin ZrO_2 films while suppressing leakage current, we fabricated interdigitated electrodes (IDEs) to facilitate in-plane (IP) polarization switching (Fig. 3D). In IDE structures, leakage is no longer limited by the ZrO_2 thickness (~5 Å to 10 nm) but rather by the IP electrode spacing (~1 μm). The expected field-induced nonpolar-to-polar phase transition for fluorite-structure antiferroelectrics (*8*, *18*), illustrated by double-switching *P-V* behavior in the IDE structures, is observed for thick (5 and 10 nm) t-ZrO_2 films (Fig. 3E and

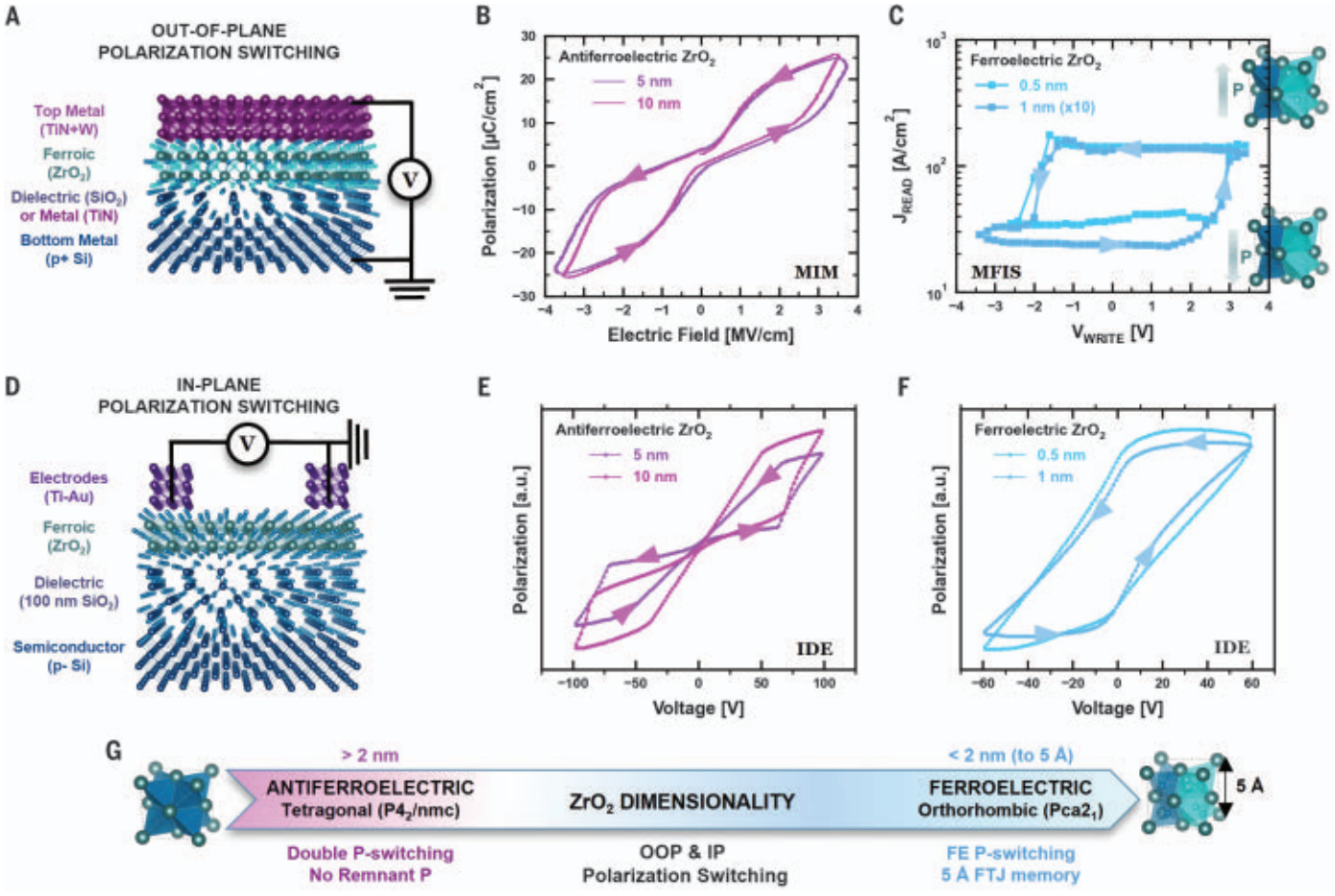


Fig. 3. Thickness-dependent polarization switching in ultrathin ZrO_2. (**A**) Schematic OOP capacitor geometries—MIM capacitors with bottom TiN and MFIS tunnel junctions with SiO_2 interlayer dielectrics—used to investigate thickness-dependent OOP polarization switching in ZrO_2. (**B**) Antiferroelectric-like OOP polarization switching observed in relatively thick (5 and 10 nm) ZrO_2 from MIM *P-V* hysteresis loops. (**C**) Ferroelectric-like OOP polarization switching observed in ultrathin (5 Å and 1 nm) ZrO_2 from pulsed *I-V* hysteresis loops measured in MFIS tunnel junctions, demonstrating two bistable remnant resistive states, consistent with ferroelectric polarization switching and piezoresponse force microscopy hysteresis loops (fig. S14). J_{READ}, read current; V_{WRITE}, write voltage. (**D**) Schematic IP device geometry (IDEs) used to investigate thickness-dependent IP polarization switching in ZrO_2. (**E** and **F**) Antiferroelectric-like (E) and ferroelectric-like (F) IP polarization switching observed in relatively thick (5 and 10 nm) ZrO_2 and ultrathin (5 Å and 1 nm) ZrO_2, respectively, from IDE *P-V* hysteresis loops. (**G**) Dimensionality-driven antiferroelectric-to-ferroelectric evolution of ZrO_2 demonstrated through OOP and IP polarization switching. FE, ferroelectric.

fig. S10). As ZrO_2 drops below the critical 2-nm thickness, *P-V* behavior for 1-nm and 5-Å ZrO_2 displays ferroelectric-like counterclockwise polarization switching (Fig. 3F and fig. S10). The polarization switching in this IP geometry is consistent with SHG imaging (Fig. 2C) and SHG spectra (fig. S11), whose geometry (fig. S12) is sensitive to IP inversion symmetry breaking (*18*) rather than out-of-plane (OOP) inversion symmetry breaking mapped by piezoresponse force microscopy imaging (fig. S13).

Pulsed current-voltage (*I-V*) measurements in metal-ferroelectric-insulator-semiconductor (MFIS) tunnel junction structures (Fig. 3C) provide additional ferroic phase insights into ZrO_2, in which tunnel electroresistance reflects the ferroelectric polarization evolution with field (*18*). For ultrathin (1 and 0.5 nm) ZrO_2 tunnel barriers (Fig. 3C), abrupt bistable resistance states exhibit counterclockwise hysteresis, consistent with ferroelectric polarization switching in the MFIS geometry and voltage polarity–independent hysteretic behavior (fig. S14). Similar tunnel electroresistance hysteresis maps have been shown for ferroelectric tunnel junctions that integrate ultrathin Zr:HfO_2 barriers (*11*, *22*, *23*), but until this point, fluorite-based ferroelectric tunnel junctions have not been demonstrated below 1 nm.

To unravel these anomalous size effects, specifically the emergent 2D ferroelectricity in a conventionally paraelectric material, surface energies, which take on an amplified role in the ultrasmall and ultrathin regime (*21*), can provide key insights into fluorite polymorphic phase stability (*15*). First-principles calculations that consider surface-energy contributions predict lower o-phase Gibbs free energy relative to the t-phase at ultrasmall sizes (<4 nm) and large pressures for the HfO_2-ZrO_2 system (*21*). Indeed, pressure-driven stabilization of the o-phase in ZrO_2 has been observed through hydrostatic pressure in bulk ZrO_2 (*15*) and epitaxial strain in thin-film ZrO_2 (*16*, *24*). Here, 2D thickness scaling—that is, confinement in the vertical dimension (Fig. 1A)—should trigger similar pressure-driven and surface energy–induced effects. Accordingly, the ferroelectric o-phase is stabilized at reduced dimensions, as expected for ultrasmall crystallite sizes (*20*) and nanoscale grain sizes (*21*) in fluorite binary oxides. On the other hand, similar size effects typically destabilize the polar crystal structure in the conventional perovskite ferroelectrics

(*25*), marking one key distinction between the two ferroelectric systems.

Besides simply stabilizing atomic-scale ferroelectricity, another puzzling feature lies in the amplified markers of polar distortion with decreased thickness (Fig. 2). Indeed, ZrO_2 demonstrates many of the same ultrathin-enhanced lattice distortion signatures as ferroelectric polycrystalline Zr:HfO_2 (*11*) and epitaxial Zr:HfO_2 (*10*, *26*) films. Therefore, these "reverse" size effects relative to its perovskite counterparts, in which polarization typically decreases with decreasing thickness (*7*), may be intrinsic to fluorites. Recent first-principles calculations indicate that the fluorite-structure o-phase ($Pca2_1$) displays a negative longitudinal piezoelectric effect (*27*), in which compression along the polarization direction leads to a larger polar distortion. Therefore, the negative longitudinal piezoelectric effect could provide an atomic-scale mechanism underlying the observed increased polar distortion as thickness is reduced to the ultrathin regime (*10*, *11*, *26*) in fluorite-structure ferroelectrics. This would mark another distinguishing feature from prototypical perovskite-based ferroelectric thin films, which demonstrate positive longitudinal piezoelectricity and diminished polar distortion at the atomic scale (*7*).

Along with the distinct piezoelectric origins (*27*), unconventional ferroelectric origins have also been attributed to fluorite-structure binary oxides (*12*, *13*). First-principles calculations suggest that 2D fluorite-structure $Pca2_1$ slabs maintain switchable polarization because of their improper nature (*12*); indeed, for improper ferroelectric transitions, the primary nonpolar structural distortion, from which the spontaneous polarization indirectly arises, is impervious to electrostatic depolarization effects (*28*). Therefore, improper fluorite ferroelectrics should display robust, switchable electric dipoles with no critical thickness (*12*), as we observed for ultrathin ZrO_2 (Fig. 3).

In addition to the depolarization-resistant nature of fluorite ferroelectricity (*12*, *13*), IP polarization can also help mitigate depolarization fields, which typically suppress OOP polarization in the ultrathin regime (*29*). Indeed, the ultrathin o-ZrO_2 films demonstrate IP polarization, as evidenced by IP inversion symmetry breaking (Fig. 2C) and polarization switching (Fig. 3F). These films exhibit predominant (111) OOP texture (Fig. 2A and fig. S3); considering that the $Pca2_1$ unit-cell polar axis lies along a principal lattice direction, o-ZrO_2 films project a substantial IP polarization. Therefore, the highly oriented nature of the ultrathin ZrO_2 films, preferentially stacked along their close-packed (111) planes favored by surface-energy considerations, can also contribute to the sustained atomic-scale polarization.

Considering that traditional 3D materials may possess unrealized spontaneous polarization, exemplified here by a conventionally paraelectric binary oxide developing ferroelectric order at reduced dimensions, simply scaling the thickness to the atomic scale offers a straightforward, yet effective, route to 2D ferroelectricity by design in intrinsically centrosymmetric materials. Therefore, reduced dimensionality offers a powerful inversion symmetry–breaking methodology (*30*), along with epitaxial strain (*31*, *32*) and twisted heterostructures (*33*), for materials in proximity to pressure-induced polar instabilities, such as other simple binary oxides (*34*).

Specifically regarding the HfO_2-ZrO_2 binary oxide family, the emergence of atomic-scale ferroelectricity in ZrO_2 underscores the distinct nature of fluorite-structure size effects, in which reduced dimensionality induces ferroelectric order even in its conventionally antiferroelectric endmember, not just Zr:HfO_2 (*11*). Therefore, thickness-scaling alone can span the fluorite ferroelectric-antiferroelectric phase diagram, moving beyond the established HfO_2-ZrO_2 composition space (*4*, *8*, *9*). Furthermore, the observed polarization switching (Fig. 3) to the fluorite-structure unit-cell size, 5 Å, validates recent predictions of its unorthodox ferroelectric origins (*12*, *13*), likely marking the thinnest demonstration of hysteretic polarization switching in any ferroelectric system (table S1). Critically, the polarization switching for 5-Å ZrO_2 persists beyond 125°C (fig. S15), which is promising for electronic applications, such as nonvolatile ferroelectric memory (Fig. 3C). Therefore, simple fluorite-structure binary oxides offer a model material system not just to explore unconventional piezoelectric and ferroelectric phenomena at the 2D limit but also for integration into highly scaled next-generation Si electronics.

REFERENCES AND NOTES

1. S. Salahuddin, K. Ni, S. Datta, *Nat. Electron.* **1**, 442–450 (2018).
2. M. Dawber, K. M. Rabe, J. F. Scott, *Rev. Mod. Phys.* **77**, 1083–1130 (2005).
3. T. S. Böscke, J. Müller, D. Bräuhaus, U. Schröder, U. Böttger, *Appl. Phys. Lett.* **99**, 102903 (2011).
4. J. Müller *et al.*, *Nano Lett.* **12**, 4318–4323 (2012).
5. T. Mikolajick, U. Schroeder, M. H. Park, *Appl. Phys. Lett.* **118**, 180402 (2021).
6. H. Qiao, C. Wang, W. S. Choi, M. H. Park, Y. Kim, *Mater. Sci. Eng. Rep.* **145**, 100622 (2021).
7. J. F. Ihlefeld *et al.*, *J. Am. Ceram. Soc.* **99**, 2537–2557 (2016).
8. T. Mikolajick *et al.*, *J. Appl. Phys.* **129**, 100901 (2021).
9. U. Schroeder *et al.*, *Jpn. J. Appl. Phys.* **58**, SL0801 (2019).
10. Y. Wei *et al.*, *Nat. Mater.* **17**, 1095–1100 (2018).
11. S. S. Cheema *et al.*, *Nature* **580**, 478–482 (2020).
12. H.-J. Lee *et al.*, *Science* **369**, 1343–1347 (2020).
13. B. Noheda, J. Íñiguez, *Science* **369**, 1300–1301 (2020).
14. M. Dogan *et al.*, *Nano Lett.* **18**, 241–246 (2018).
15. O. Ohtaka *et al.*, *Phys. Rev. B* **63**, 174108 (2001).
16. T. Song *et al.*, *ACS Appl. Electron. Mater.* **3**, 2106–2113 (2021).
17. C. A. Randall, Z. Fan, I. Reaney, L. Chen, S. Trolier-McKinstry, *J. Am. Ceram. Soc.* **104**, 3775–3810 (2021).
18. Materials, methods, and additional information are available as supplementary materials.
19. P. Nukala *et al.*, *Science* **372**, 630–635 (2021).
20. H. Du *et al.*, *Matter* **4**, 986–1000 (2021).
21. R. Materlik, C. Künneth, A. Kersch, *J. Appl. Phys.* **117**, 134109 (2015).
22. Y. Wei *et al.*, *Phys. Rev. Appl.* **12**, 031001 (2019).
23. S. S. Cheema *et al.*, *Adv. Electron. Mater.* 10.1002/aelm.202100499 (2021).
24. S. E. Reyes-Lillo, K. F. Garrity, K. M. Rabe, *Phys. Rev. B* **90**, 140103 (2014).
25. W. J. Merz, *Phys. Rev.* **78**, 52–54 (1950).
26. J. Lyu, I. Fina, R. Solanas, J. Fontcuberta, F. Sánchez, *ACS Appl. Electron. Mater.* **1**, 220–228 (2019).
27. J. Liu, S. Liu, J.-Y. Yang, L. Liu, *Phys. Rev. Lett.* **125**, 197601 (2020).
28. N. Sai, C. J. Fennie, A. A. Demkov, *Phys. Rev. Lett.* **102**, 107601 (2009).
29. J. Junquera, P. Ghosez, *Nature* **422**, 506–509 (2003).
30. D. Lee *et al.*, *Science* **349**, 1314–1317 (2015).
31. M. P. Warusawithana *et al.*, *Science* **324**, 367–370 (2009).
32. J. H. Haeni *et al.*, *Nature* **430**, 758–761 (2004).
33. E. Tsymbal, *Science* **372**, 1389–1390 (2021).
34. R. Batra *et al.*, *Chem. Mater.* **32**, 3823–3832 (2020).

ACKNOWLEDGMENTS

S.S.C. and N.S. thank C. Tassone, A. Mehta, and K. Stone for experimental support at the Stanford Synchrotron Radiation Lightsource (SSRL) and R. Ramesh for access to scanning probe microscopy. **Funding:** This research was supported in part by the following: the Berkeley Center for Negative Capacitance Transistors (BCNCT); Applications and Systems-Driven Center for Energy-Efficient Integrated NanoTechnologies (ASCENT), one of the six centers in the Joint University Microelectronics Program (JUMP) initiative, a Semiconductor Research Corporation (SRC) program sponsored by Defense Advanced Research Projects Agency (DARPA); the DARPA Foundations Required for Novel Compute (FRANC) program; the Interaction of Ionizing Radiation with Matter (IIRM) University Research Alliance supported by the Department of Defense; and the US Department of Energy (DOE), Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division, under contract no. DE-AC02-05-CH11231 (Microelectronics Co-Design program) for the development of materials for low-power microelectronics. This research used resources of the Advanced Photon Source (APS), a DOE Office of Science user facility operated for the DOE Office of Science by Argonne National Laboratory under contract no. DE-AC02-06CH11357. V.A.S. and J.W.F. were supported by the DOE, Office of Science, Basic Energy Sciences, under award no. DE-SC-0012375. Use of the SSRL, SLAC National Accelerator Laboratory, is supported by the DOE, Office of Science, Office of Basic Energy Sciences, under contract no. DE-AC0276SF00515. This research used resources of the Advanced Light Source (ALS), which is a DOE Office of Science User Facility under contract no. DE-AC02-05CH11231. Electron microscopy was performed at the Molecular Foundry, Lawrence Berkeley National Laboratory, supported by the DOE, Office of Science, Office of Basic Energy Sciences (DE-AC02-05CH11231). **Author contributions:** S.S.C. conceived the project and designed the research and experiments. S.S.C. performed film synthesis and ferroic phase optimization. N.S. performed device fabrication. N.S. and S.S.C. performed dielectric and electrical measurements. N.S. performed scanning probe microscopy with the supervision of J.C. S.-L.H. performed transmission electron microscopy and analysis. Y.R. performed second-harmonic generation with the supervision of C.P.G. S.S.C., C.-H.H., P.S., V.A.S., and J.W.F. performed synchrotron spectroscopy (at ALS and APS). S.S.C., N.S., V.A.S., and Z.Z. performed synchrotron diffraction (at SSRL and APS). S.S.C. wrote the manuscript. S.S.C., N.S., and S.S. edited the manuscript. S.S. supervised the research. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data are available in the manuscript or the supplementary materials.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abm8642
Materials and Methods
Supplementary Text
Figs. S1 to S15
Table S1
References (*35*–*132*)

Submitted 18 October 2021; resubmitted 14 February 2022
Accepted 1 April 2022
10.1126/science.abm8642

PALEOECOLOGY

A stronger role for long-term moisture change than for CO_2 in determining tropical woody vegetation change

William D. Gosling[1]*, Charlotte S. Miller[2], Timothy M. Shanahan[3], Philip B. Holden[4], Jonathan T. Overpeck[5], Frank van Langevelde[6,7]

Anthropogenically elevated CO_2 (eCO_2) concentrations have been suggested to increase woody cover within tropical ecosystems through fertilization. The effect of eCO_2 is built into Earth system models, although testing the relationship over long periods remains challenging. Here, we explore the relative importance of six drivers of vegetation change in western Africa over the past ~500,000 years (moisture availability, fire activity, mammalian herbivore density, temperature, temperature seasonality, CO_2) by coupling past environmental change data from Lake Bosumtwi (Ghana) with global data. We found that moisture availability and fire activity were the most important factors in determining woody cover, whereas the effect of CO_2 was small. Our findings suggest that the role of eCO_2 effects on tropical vegetation in predictive models must be reconsidered.

Tropical vegetation forms an important part of the global carbon cycle, with forests contributing 33% (*1*) and savannas 30% (*2*) of the global terrestrial net primary productivity. The current paradigm suggests that anthropogenically elevated CO_2 (eCO_2) levels have enhanced carbon sequestration rates in tropical forests (*3*, *4*) and resulted in elevated woody cover, "greening," within tropical savannas and grasslands (*5*–*7*). The suggested reason for these changes is that plant species that use the C_3 photosynthetic pathway (such as trees and shrubs) sequester more carbon under eCO_2 than species that use the C_4 pathway (such as tropical grasses) (*8*, *9*). This highly influential concept of tropical greening caused by eCO_2 is difficult to test, especially on large spatial scales or over long periods of time (*10*–*12*). However, dynamic global vegetation and Earth system models, which are used to predict the impacts of projected future climate change, nevertheless typically include a fertilization effect of eCO_2 on the growth of woody cover in the tropics [e.g., (*6*, *13*–*16*)].

The impact of eCO_2 on vegetation has recently been challenged in observational and experimental field studies running over decades (*17*–*20*); with atmospheric CO_2 rising by ~100 ppmv (parts per million by volume) since 1960 (*21*). Therefore, a pertinent question now exists regarding the effect of eCO_2 on vegetation community change (e.g., the shift in the dominance of plants with C_3 versus C_4 photosynthetic pathways) rather than on individual plants, and therefore on the importance of including eCO_2 as a driver of vegetation change in models over long periods of time and on large spatial scales. This is fundamental to understanding Earth system function because it took thousands of years for modern tropical forest vegetation to establish after the end of the last global glacial period [e.g., (*22*)]. Furthermore, if we are to be able to correctly anticipate future vegetation change and predict the potentially significant roles that vegetation could play in mitigating, or enhancing, climate change under continuing anthropogenic eCO_2 levels, it is over long periods of time that knowledge is required. Although modeling studies have been used to assess how lower eCO_2 levels could have affected vegetation during the latter part of the last glacial period, ~21,000 to 18,000 years ago (*23*–*25*), there have been no data available to support the causal relationship between eCO_2 and tropical woody cover over multiple high magnitude (>100 ppmv) eCO_2 transitions relevant to the establishment of tropical vegetation communities. Here, we remedy this problem and explore the relationship among tropical woody cover, CO_2, and five other variables known to influence woody cover [fire activity, mammalian herbivore density, moisture availability, temperature, and temperature seasonality (*26*–*28*)], over the past ~500,000 years at a site in tropical western Africa.

The sedimentary archive recovered from Lake Bosumtwi in Ghana (6°30'N, 1°25'W, 97 m above sea level) contains evidence for multiple multimillennial transitions between tropical woody and grass-dominated vegetation over the past ~500,000 years (*29*, *30*). Located within moist, semi-evergreen forest today, Bosumtwi receives ~1450 mm of rainfall per year (*31*). To the north of Bosumtwi, there is a concomitant decline in rainfall and woody cover (*32*). The relationship between the depth and age of the sediments recovered from Bosumtwi is constrained independently by radiocarbon, optically stimulated luminescence, and uranium-thorium ages. All of these data were combined in a Bayesian probability model to allow uncertainties in the chronology to be calculated (fig. S1) (*33*). Its exceptionally long record of past environmental changes and its position on ecological and climatic gradients make Bosumtwi the ideal location to test the relative importance of drivers of tropical vegetation dynamics (*22*). Furthermore, the variation in woody cover in the Bosumtwi record and the simulated vegetation carbon storage show similar, sub-100,000 year patterns of change over the past ~500,000 years (*34*) (Fig. 1, A and B).

We characterized the woody cover in tropical western Africa, and the likely factors driving change, using a combination of observed and simulated data obtained from the sedimentary core recovered from Bosumtwi (*35*), the GENIE-1 Earth system (climate-carbon cycle) model (*36*), and ice core records (*37*). Four variables were obtained from Bosumtwi, pollen taxa, charcoal fragments, *Sporormiella* fungal spores, and $\delta^{15}N$, providing time series of vegetation cover (variation of woody versus grassy cover), fire activity, mammalian herbivore density, and moisture availability, respectively (*33*). Two additional variables were obtained for the Bosumtwi region from the GENIE-1 model: temperature and temperature seasonality (*34*). Atmospheric CO_2 concentrations were taken from ice-core records (*37*). The datasets were integrated using an updated chronology for Bosumtwi (*33*), the chronology from within the GENIE-1 model (*36*), and the ice core chronology (*37*).

The data extracted from the Bosumtwi sediment core show variations in the degree of woody cover (Fig. 1B), fire activity (Fig. 1C), mammalian herbivore density (Fig. 1D), and moisture availability (Fig. 1E). The major variation in the pollen assemblage over the past ~500,000 years is driven primarily by the abundance of grass (Poaceae) (Fig. 1B) (*29*, *30*, *33*). Fire activity and mammalian herbivores are shown to have been near continually present in the landscape for the past ~500,000 years, with the lowest prevalence of fire activity being after ~9000 years ago and the lowest density of mammalian herbivores between ~22,000 and 800 years ago. Moisture availability is shown to oscillate throughout the past ~500,000 years, with lower-than-average availability suggested between ~300,000 and 130,000 years ago and ~30,000 years ago (Fig. 1E). Temperature (Fig. 1F), temperature seasonality (Fig. 1G), and atmospheric CO_2 levels (Fig. 1H) show major fluctuations on 100,000-year and 21,000-year time scales related to orbital climate forcing (*34*).

[1]Institute for Biodiversity & Ecosystem Dynamics, University of Amsterdam, Netherlands. [2]Leeds Trinity University, Leeds, UK. [3]Department of Geological Sciences, University of Texas at Austin, Austin, TX, USA. [4]School of Environment, Earth & Ecosystem Sciences, The Open University, Milton Keynes, UK. [5]School for Environment & Sustainability, University of Michigan, Ann Arbor, MI, USA. [6]Department of Environmental Sciences, Wageningen University & Research, Wageningen, Netherlands. [7]School of Life Sciences, University of KwaZulu-Natal, KwaZulu-Natal, South Africa.
*Corresonding author. Email: w.d.gosling@uva.nl

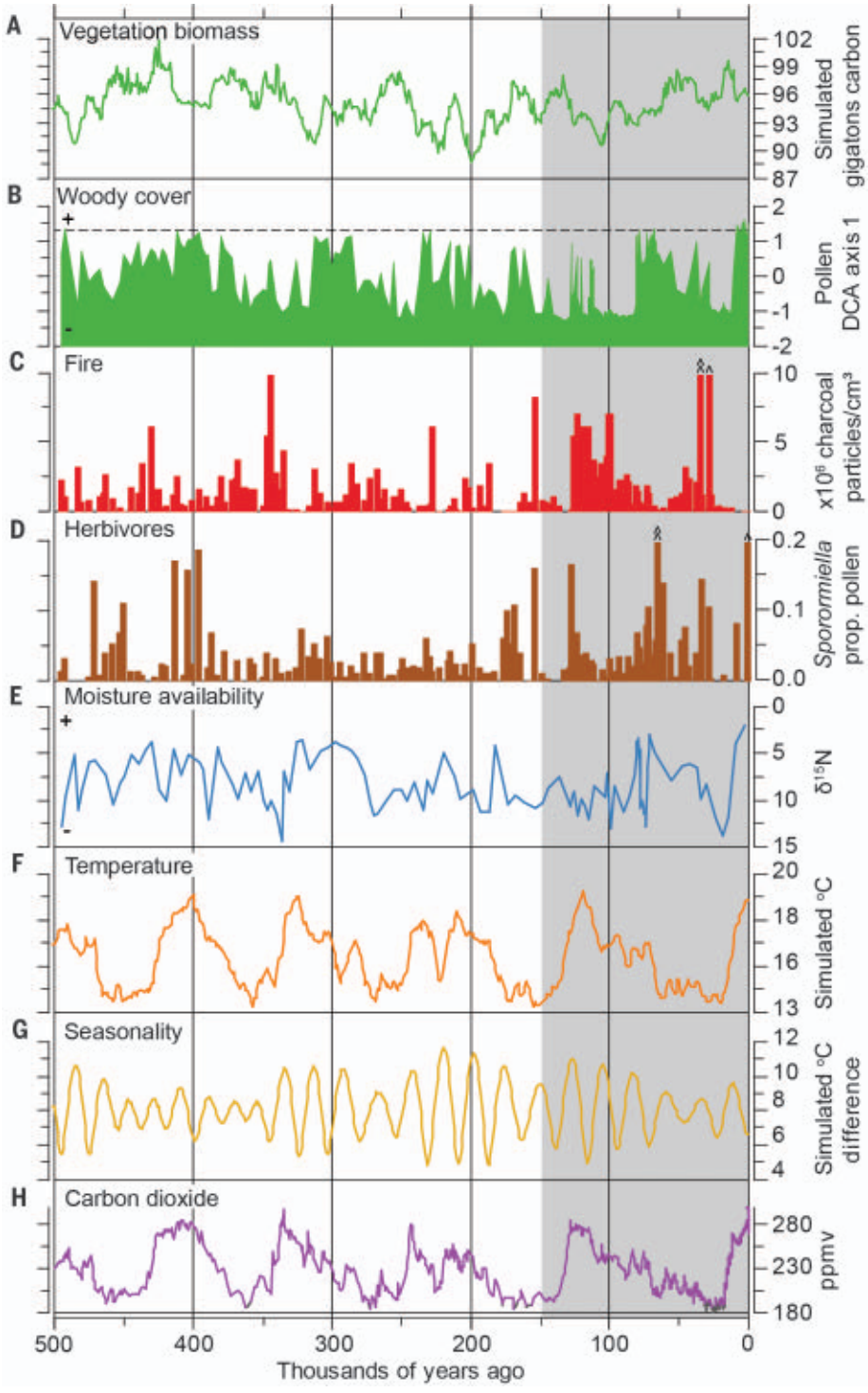


Fig. 1. Five hundred thousand years of tropical environmental change. Shown are model simulations of conditions for low northern latitudes (0 to 23°N) [(A), (F), and (G)], empirical data from Lake Bosumtwi (Ghana) [(B) to (E)], and global atmospheric CO_2 (H). (**A**) Simulated vegetation carbon storage. (**B**) Characterization of the variation in the abundance of woody taxa within the overall pollen/spore assemblage data; plus sign indicates more woody taxa and minus sign indicates higher abundance of grass (represented by the first ordination axis that is based on detrended correspondence analysis: DCA axis 1). (**C**) Reconstructed fire activity from sedimentary charcoal abundance data; two values exceed the scales shown, sample indicated by ^^ = 25 and ^ = 16 × 10^6 charcoal particles/cm^3. (**D**) Relative mammalian herbivore density reconstructed from sedimentary *Sporormiella* abundance; two samples exceed the scale shown, sample indicated by ^^ = 0.37 and ^ = 0.36 as a proportion of the pollen sum. (**E**) Moisture availability reconstructed from sedimentary $\delta^{15}N$; plus sign indicates deeper lake levels and minus sign indicates lower lake levels. (**F**) Simulated temperature for low northern latitude band. (**G**) Simulated difference in temperature between seasons for low northern latitude band; high values indicate greater difference in temperature between seasons. (**H**) Atmospheric CO_2 [for further information, see (*33*)]. Gray box indicates data used within the ~150,000-year time interval SEM (Fig. 2B). The entire time range of the figure indicates the data used within the ~500,000-year time interval SEM (Fig. 2C).

The direction and strength of the relationships between the various variables were assessed through structural equation modeling (SEM) [e.g., (*26*)]. SEM identifies the directionality of linear relationships within an a priori model of seven variables (Fig. 2A) (*33*) calculated independently over two different intervals of time (~150,000 to 0 years and 500,000 to 0 years). The ~150,000-year interval was selected because this section of the sedimentary archive contains the largest number of independent radiometric age control points; i.e., the chronological control is the most robust through this section of the record. The ~500,000-year interval was used because it includes the largest amount of available data on past ecological and climatic change (pollen, *Sporormiella* fungal spores, and charcoal fragments with n = 112; $\delta^{15}N$ with n = 51). The SEM was constructed to examine hypothesized direct and indirect relationships based on current understanding of tropical vegetation (*26*, *32*) (Fig. 2A). To explore for possible bias caused by chronological uncertainty in the Bosumtwi record, the SEMs were also run against 1000 possible alternative versions of the chronology (*29*) for both intervals of time [for the ~500,000-year data, see Fig. 3; other analysis is in (*33*)]. Given the limitations of null-hypothesis significance testing (*38*), we focused on the relative importance

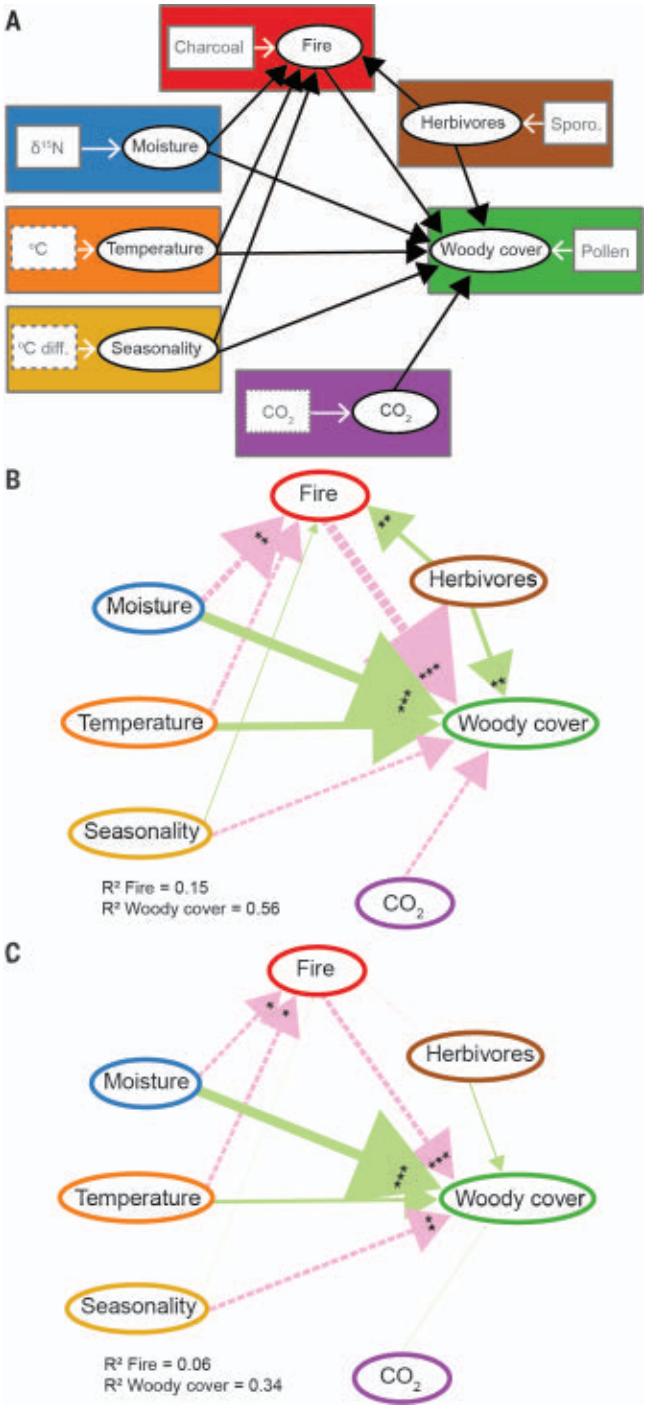


Fig. 2. Assessment of past relationships among climate, atmospheric CO_2, landscape processes, and vegetation in the tropics using a SEM approach. (**A**) Hypothesized relationships examined in this study (see Fig. 1). (**B** and **C**) Results of SEM for the past ~150,000 years (B) and the past ~500,000 years (C) based the weighted average chronology (*33*). Measured variables are indicated by square gray boxes: solid lines indicate empirical data derived from the Bosumtwi sedimentary record, dashed lines indicate data derived from the GENIE-1 model, and dotted lines indicate empirical data from ice cores [see (*33*)]. Latent variables are in black ovals. Colors associated with variables match Fig. 1. In (B) and (C), arrow color/form indicates positive (green, solid) and negative (pink, dashed) relationships, and arrow thickness represents the absolute strength of the relationships based on the average age model (*33*). Significance testing: *0.01 < P < 0.05; **0.001 < P < 0.01; and ***P < 0.001. Full model results are presented in table S1 (*33*).

of the predictor variables using the distribution of the standardized coefficients of the SEMs.

Regardless of time interval, moisture availability is shown to be the most important driver of woody cover, directly and indirectly, on multimillennial time (Figs. 2, B and C, and 3 and table S1). Increasing moisture availability simultaneously results in increasing woody cover and a decrease in grass-fueled fire activity. This decrease likely reflects changes in the abundance of the woody component of these ecosystems, resulting in the negative relationships between fire activity and woody cover (*39*). Mammalian herbivore density shows a strong positive relationship with fire activity and woody cover. These findings are in agreement with recent analyses indicating that herbivore biomass and woody cover are positively correlated when woody cover is sparse (*27*). Conversely, increasing temperature seasonality had a strong negative effect on woody cover. These findings are consistent with those based on recent global observational studies in savannas (*26*, *28*, *32*) and forests (*40*), and additionally suggests that the mechanistic controls on tropical woody cover have remained stable for at least the past half a million years. Unexpectedly, the effect of atmospheric CO_2 on woody cover both directly and indirectly, through temperature, was small during the past ~500,000 years at Bosumtwi (Fig. 2, B and C). Furthermore, even when temperature or moisture were excluded from the SEM, the relative importance of the different drivers was found to hold [presented in (*33*)].

Our findings suggest that abiotic (moisture availability and fire activity) and biotic (mammalian herbivore density) drivers operating at the landscape scale, together with temperature seasonailty, can override the role of CO_2 in driving woody growth at forest-savanna transitions in the tropics. The absence of a strong relationship between CO_2 and woody cover at Bosumtwi provides evidence against the idea that lower eCO_2 levels could be mainly responsible for tropical vegetation change over glacial-interglacial cycles (*23*), and supports findings of a stronger role for precipitation than CO_2 in determining the position of the treeline in African mountain regions (*24*, *25*). Comparison of major vegetation shifts identified in other long terrestrial records from Africa reveals no strong and contrasting correlation with each other and shifts in global CO_2 (figs. S7 and S8) (*33*). In eastern Africa during the past ~500,000 years, the periods of highest woody cover (i.e., the lowest abundance of grasses) at Lake Magadi and Lake Malawi are antiphased (*41*–*43*), suggesting that a regionally specific change [such as in precipitation pattern sensu (*44*)], rather than a uniform CO_2 fertilization effect, is likely to be the driving factor. Shorter records from southern [Vankervelsvlei (*45*)] and central [Lake Bambili (*46*)] Africa show changes in woody cover during the last glacial cycle (the past ~100,000 years) that do not all occur in concert with global CO_2 shifts, again suggesting that other processes were important. A recent analysis of African climate change over the past 1 million years revealed that the major change in precipitation patterns occurred ~300,000 years ago and was related to a shift in tropical climate systems (Walker circulation) (*44*, *47*). On the basis of the strong control of moisture availability on woody cover shown here, we hypothesize that changes in moisture availability are likely to exerted a greater control on vegetation across Africa than CO_2.

The low relative importance of CO_2 in determining woody cover in Africa over long time scales can be considered mechanistically. Although trees require more carbon to deploy a unit of leaf area than grasses (*5*), and they need to allocate large amounts of carbon in roots to resprout after grass fire and herbivory damage (*48*, *49*) and to ensure that they rapidly attain the sizes needed to prevent this damage (*50*),

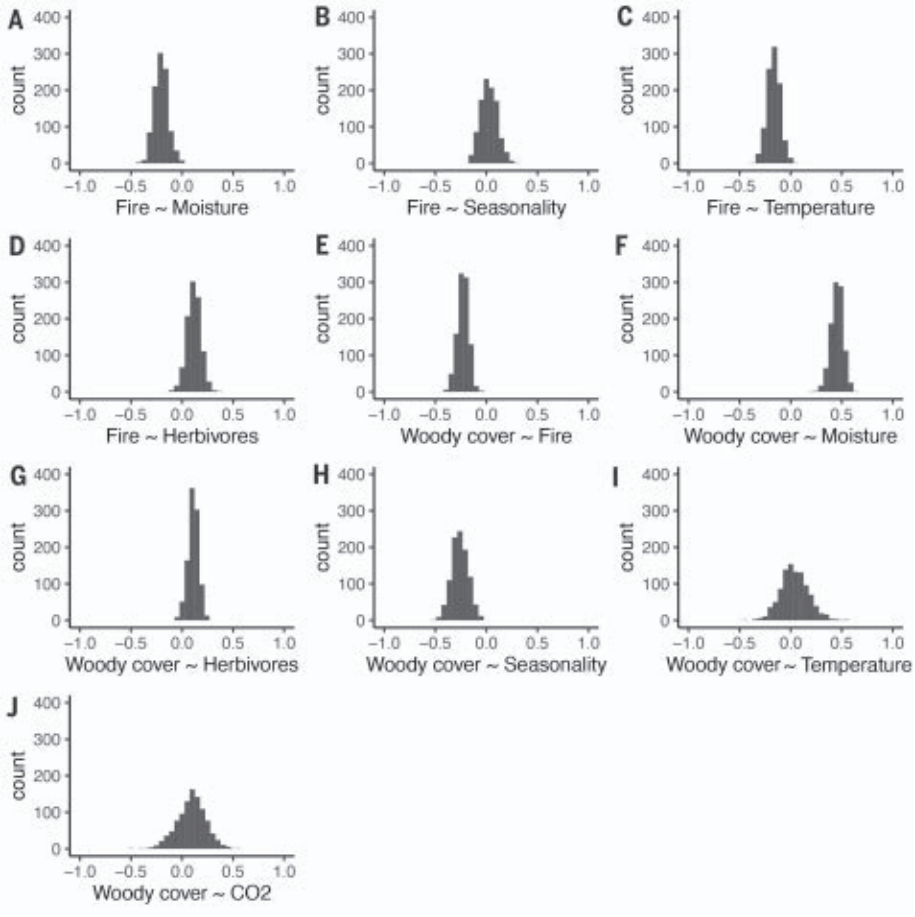


Fig. 3. Distributions of the standardized coefficients for the predictor variables in the SEMs for the ~500,000-year interval for Bosumtwi (Fig. 2C). The SEMs were run against 1000 possible alternative versions of the chronology to explore for possible bias caused by chronological uncertainty. The predictor variables are shown per response variable (fire and woody cover). The most important drivers of woody cover are moisture availability and fire because they do not include zero in their distributions. The statistics are reported in (*33*).

eCO2 may not necessarily result in higher woody cover. Possible explanations for the small CO_2 effect on woody cover are that it can be counteracted by impacts from climate-driven changes (*51*) or by water or nutrient limitations on vegetation productivity (*17*), or fire could limit woody cover expansion in tropical grasslands and savannas (*22*) [but see (*28*)]. Generating comparable, high-resolution, multiproxy records to test our findings in different biomes across the tropics is now a priority.

Our study confirms that an improved understanding of vegetation community change (C_4 versus C_3 plant response) under eCO_2 is needed for dynamic global vegetation and Earth system modeling of the future. The small effects of CO_2 on past woody cover challenges the long-term capacity of tropical savannas and forests to slow anthropogenically eCO_2 through increased greening as woody taxa recruitment and growth are enhanced through CO_2 fertilization effects, as predicted by these models (*14*–*16*). Therefore, our findings suggest that the long-term effectiveness of proposed tropical vegetation-based strategies to mitigate anthropogenic eCO_2 through enhanced carbon sequestration rates (*52*, *53*) is contingent on sufficient moisture availability.

REFERENCES AND NOTES

1. R. A. Houghton, *Glob. Change Biol.* **11**, 945–958 (2005).
2. J. Grace, J. S. José, P. Meir, H. S. Miranda, R. A. Montes, *J. Biogeogr.* **33**, 387–400 (2006).
3. O. L. Phillips *et al.*, *Science* **282**, 439–442 (1998).
4. S. L. Lewis, J. Lloyd, S. Sitch, E. T. Mitchard, W. F. Laurance, *Annu. Rev. Ecol. Evol. Syst.* **40**, 529–549 (2009).
5. W. J. Bond, G. F. Midgley, *Glob. Change Biol.* **6**, 865–869 (2000).
6. S. I. Higgins, S. Scheiter, *Nature* **488**, 209–212 (2012).
7. Z. Zhu *et al.*, *Nat. Clim. Chang.* **6**, 791–795 (2016).
8. E. A. Ainsworth, S. P. Long, *New Phytol.* **165**, 351–371 (2005).
9. J. R. Ehleringer, T. E. Cerling, B. R. Helliker, *Oecologia* **112**, 285–299 (1997).
10. D. D. Breshears *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15144–15148 (2005).
11. H. D. Adams *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 7063–7066 (2009).
12. R. J. Norby, D. R. Zak, *Annu. Rev. Ecol. Evol. Syst.* **42**, 181–203 (2011).
13. S. P. Harrison, C. I. Prentice, *Glob. Change Biol.* **9**, 983–1004 (2003).
14. S. Sitch *et al.*, *Glob. Change Biol.* **14**, 2015–2039 (2008).
15. P. M. Cox *et al.*, *Nature* **494**, 341–344 (2013).
16. C. Huntingford *et al.*, *Nat. Geosci.* **6**, 268–273 (2013).
17. S. Wang *et al.*, *Science* **370**, 1295–1300 (2020).
18. P. Groenendijk *et al.*, *Glob. Chang, Biol.* **21**, 3762–3776 (2015).
19. P. van der Sleen *et al.*, *Nat. Geosci.* **8**, 24–28 (2015).
20. P. B. Reich, S. E. Hobbie, T. D. Lee, M. A. Pastore, *Science* **360**, 317–320 (2018).
21. J. Blunden, T. Boyer, *Bull. Am. Meteorol. Soc.* **102**, S1–S475 (2021).
22. T. M. Shanahan *et al.*, *Sci. Rep.* **6**, 29587 (2016).
23. D. Jolly, A. Haxeltine, *Science* **276**, 786–788 (1997).
24. H. Wu, J. Guiot, S. Brewer, Z. Guo, *Clim. Dyn.* **29**, 211–229 (2007).
25. K. Izumi, A. Lézine, *Quat. Sci. Rev.* **152**, 93–103 (2016).
26. C. E. R. Lehmann *et al.*, *Science* **343**, 548–552 (2014).
27. G. P. Hempson, S. Archibald, W. J. Bond, *Science* **350**, 1056–1061 (2015).
28. E. M. Veenendaal *et al.*, *New Phytol.* **218**, 153–166 (2018).
29. C. S. Miller, W. D. Gosling, *Quat. Sci. Rev.* **84**, 7–25 (2014).
30. C. S. Miller, W. D. Gosling, D. B. Kemp, A. L. Coe, I. Gilmour, *J. Quaternary Sci.* **31**, 671–677 (2016).
31. T. M. Shanahan, J. T. Overpeck, W. E. Sharp, C. A. Scholz, J. A. Arko, *Hydrol. Processes* **21**, 1678–1691 (2007).
32. M. Sankaran *et al.*, *Nature* **438**, 846–849 (2005).
33. See the supplementary materials.
34. W. D. Gosling, P. B. Holden, *J. Quaternary Sci.* **26**, 463–467 (2011).
35. C. Koeberl *et al.*, *Sci. Drill.* **1**, 23–27 (2005).
36. P. B. Holden *et al.*, *Clim. Past* **6**, 431–443 (2010).
37. D. Lüthi *et al.*, *Nature* **453**, 379–382 (2008).
38. D. R. Anderson, K. P. Burnham, W. L. Thompson, *J. Wildl. Manage.* **64**, 912–923 (2000).
39. F. Van Langevelde *et al.*, *Ecology* **84**, 337–350 (2003).
40. O. L. Phillips *et al.*, *Science* **323**, 1344–1347 (2009).
41. V. M. Muiruri *et al.*, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **567**, 110247 (2021).
42. R. B. Owen *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 11174–11179 (2018).
43. S. J. Ivory, A. M. Lézine, A. Vincens, A. S. Cohen, *Glob. Chang. Biol.* **24**, 2939–2951 (2018).
44. S. Kaboth-Bahr *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2018277118 (2021).
45. L. J. Quick *et al.*, *Quat. Int.* **404**, 136–149 (2016).
46. A. M. Lézine, K. Izumi, M. Kageyama, G. Achoundong, *Science* **363**, 177–181 (2019).
47. W. D. Gosling, E. M. L. Scerri, S. Kaboth-Bahr, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **377**, 20200483 (2022).
48. K. W. Tomlinson *et al.*, *J. Ecol.* **100**, 1113–1121 (2012).
49. C. C. F. Boonman *et al.*, *New Phytol.* **227**, 156–167 (2020).
50. A. M. Hoffman, J. A. Bushey, T. W. Ocheltree, M. D. Smith, *New Phytol.* **227**, 352–364 (2020).
51. J. Liu *et al.*, *Science* **358**, eaam5690 (2017).
52. M. J. P. Sullivan *et al.*, *Science* **368**, 869–874 (2020).
53. J. F. Bastin *et al.*, *Science* **365**, 76–79 (2019).
54. Data for: W. D. Gosling, C. S. Miller, T. M. Shanahan, P. B. Holden, J. T. Overpeck, F. van Langevelde, A stronger role for long-term moisture change than for CO_2 in determining tropical woody vegetation change, Figshare (2022); https://doi.org/10.6084/m9.figshare.18319466.

ACKNOWLEDGMENTS

We thank four anonymous reviewers and the editor for their extensive comments, which have no doubt helped to improve the quality of this manuscript. **Funding:** The past environmental change data from Lake Bosumtwi used in this study were funded by a NERC/The Open University Charter studentship (NE/H525054/1 to C.S.M.) and a NERC New Investigator Award (NE/G000824/1 to W.D.G.). We thank the Lake Bosumtwi Drilling Project team, the International Continental Drilling Program, the US National Science Foundation, and National lacustrine core repository (LacCore) USA for supporting the acquisition and curation of Bosumtwi sediment cores and age model data. **Author contributions:** The study was conceived and developed by W.D.G. and F.v.L. The past environmental change data from Bosumtwi was generated by C.S.M. and W.D.G. The statistical analysis was conducted by F.v.L. Expertise on past climate change and the Bosumwti sedimentary record was provided by T.M.S. and J.T.O. Expertise on Earth system modeling was provided by P.B.H. The manuscript was written by W.D.G. and F.v.L. with contributions from all authors. **Competing interests:** The authors declare no competing interests. **Data and materials availability:** The past environmental change data reported in this manuscript are archived in Figshare (*54*).

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abg4618
Materials and Methods
Figs. S1 to S8
Table S1
References (*55*–*93*)

Submitted 12 March 2021; accepted 29 March 2022
10.1126/science.abg4618



One-Step ELISA Experiments

AMS Biotechnology announces the launch of the QuicKey Pro range of ELISA kits—easy-to-use assays that require just one step of washing and one step of sample addition, so an entire experiment can be completed in only 90 min. The total procedure time using a QuicKey Pro ELISA kit is reduced by 1–2 h from that needed with traditional ELISA kits, resulting in significant gains in efficiency. Based upon proprietary competitive and sandwich technologies, the kits allow researchers to choose the optimal ELISA format to measure their specific analyte of interest in a wide range of complex biological samples. They are designed to provide sensitive and specific in vitro quantitative determination of specific analyte concentrations, and when unopened, offer the convenience of storage at 4°C for an extended period.

AMS Biotechnology

For info: +1-617-945-5033

www.amsbio.com/quickey-pro-elisa-kits

Fluorescent-Labeled Recombinant Proteins

R&D Systems Fluorokines are designed to simplify detection of a specific chimeric antigen receptor (CAR) on CAR T cells or other cell types of interest. The fluorophores are conjugated to the bioactive protein via amine labeling, and the resulting Fluorokines are rigorously tested to ensure consistent labeling of each lot. The manufacturing process is controlled to reduce lot-to-lot variability and yields a consistent fluorochrome/protein (F/P) ratio. Finally, each protein is quality-control tested by flow cytometry. The main advantage to Fluorokines is that target cells expressing the corresponding CAR can be directly stained and detected by flow cytometry. This method of evaluating CAR expression is highly specific, reduces processing time, and eliminates background staining that may occur by indirect detection of the CAR using an epitope-tagged target antigen and a fluorophore-labeled secondary antibody.

R&D Systems

For info: +1-800-343-7475

www.rndsystems.com

Anti-Pertuzumab Antibodies

Bio-Rad Laboratories introduces four inhibitory antibodies that are highly specific to pertuzumab (Perjeta). These ready-made antibodies inhibit the binding of the drug to its target, human epidermal growth factor receptor 2 (HER2), enabling researchers to develop highly selective, sensitive assays for bioanalysis and drug monitoring of pertuzumab and its biosimilars. Pertuzumab is a targeted cancer drug that binds to the HER2 protein, which is overexpressed in HER2-positive cancers, preventing cell growth and proliferation. The new range of anti-pertuzumab antibodies includes monovalent fragment antigen-binding (Fab)-format antibodies suitable for development of a pharmacokinetic bridging ELISA to measure free drug concentrations, as well as fully human immunoglobulin G1 (IgG1) antibodies with varying levels of affinity, which can be used as a surrogate positive control or reference standard in an antidrug antibody assay. The antibodies are generated using the Human Combinatorial Antibody Libraries (HuCAL) and CysDisplay, a proprietary method of phage display, along with guided selection methods to obtain highly targeted reagents. The recombinant production method ensures a consistent and secure supply.

Bio-Rad Laboratories

For info: +1-800-424-6723

www.bio-rad-antibodies.com

Centrifuges

Eppendorf expanded its centrifuge offerings with the CP22N and CP30NX high-speed, floor-standing centrifuges in 2021, which have now been completed with the addition of the CP-NX and CS-(F) NX series ultracentrifuges this year. These new models enrich the Eppendorf portfolio with comprehensive solutions for customers working in fields such as molecular biology, biochemistry, and cell biology. The extended product line, which includes rotors, adapters, consumables, and services, enables Eppendorf to offer centrifugation solutions for an even broader range of applications at speeds up to 150,000 rpm (1,050,000 × g). Eppendorf's new premium ultracentrifuges and micro-ultracentrifuges increase its commitment to produce high-quality products equipped for optimal performance and ease-of-use. Their ergonomic design—including a large touchscreen, fast startup time, compact design, and low noise level—gives you quick and comfortable operation. The unique rotor-life management system of the CP-NX series provides convenient tracking of rotor usage, eliminating the need for manual logbooks and extending the overall rotor lifetime. Smart, noncontact imbalance sensors safeguard users during ultraspeed applications and allow for quick sample balancing by eye (within 5-mm sample-height difference), with no need for scales.

Eppendorf

For info: +1-800-645-3050

www.eppendorf.com/your-centrifuge-solution

Peptide Synthesizer

Gyros Protein Technologies has launched PurePep Sonata+, a mid-pilot scale peptide synthesizer. PurePep Sonata+ builds on the ability of Sonata XT through the introduction of real-time flow technology, a noninvasive method to continually monitor the flow rate of fluids within the system, enabling precise fluid transfers for a range of viscosities. The need to calibrate is thereby minimized, and workflows that previously required one day for fluid calibration are significantly shortened. Our proprietary valve block system, the PurePep Pathway, is also integrated to eliminate cross-contamination, ensuring peptide purity. Thanks to reduced calibration time to streamline and shorten workflows, combined with improved reliability and robustness, this peptide synthesizer is easy to operate, and the intuitive software platform—compliant with 21 CFR Part 11—facilitates a smooth transition from the bench-scale PurePep Chorus to mid-pilot scale.

Gyros Protein Technologies

For info: +1-877-433-9400

www.gyrosproteintechnologies.com

Electronically submit your new product description or product literature information! Go to www.science.org/about/new-products-section for more information.

Newly offered instrumentation, apparatus, and laboratory materials of interest to researchers in all disciplines in academic, industrial, and governmental organizations are featured in this space. Emphasis is given to purpose, chief characteristics, and availability of products and materials. Endorsement by *Science* or AAAS of any products or materials mentioned is not implied. Additional information may be obtained from the manufacturer or supplier.

ScienceCareers online @sciencecareers.org

Senior Faculty in Cell Stress Biology
Associate or Full Members (Professors)

Roswell Park Comprehensive Cancer Center (RPCCC), Buffalo, NY / Competitive salary and startup package (depending on qualifications and rank) with full benefits

Prospective candidates must have a strong record of independent, extramurally funded program of cancer-related research that will complement the ongoing Cell Stress and Biophysical Therapies Program. Candidates are expected to have active programs in molecular, cellular and systemic mechanisms of organismal response to stresses, pathologies associated with cancer treatment and aging, as well as development of new original cancer treatment approaches. Proven ability to acquire extramural funding, along with a strong publication record and other indications of academic activity and visibility are required.

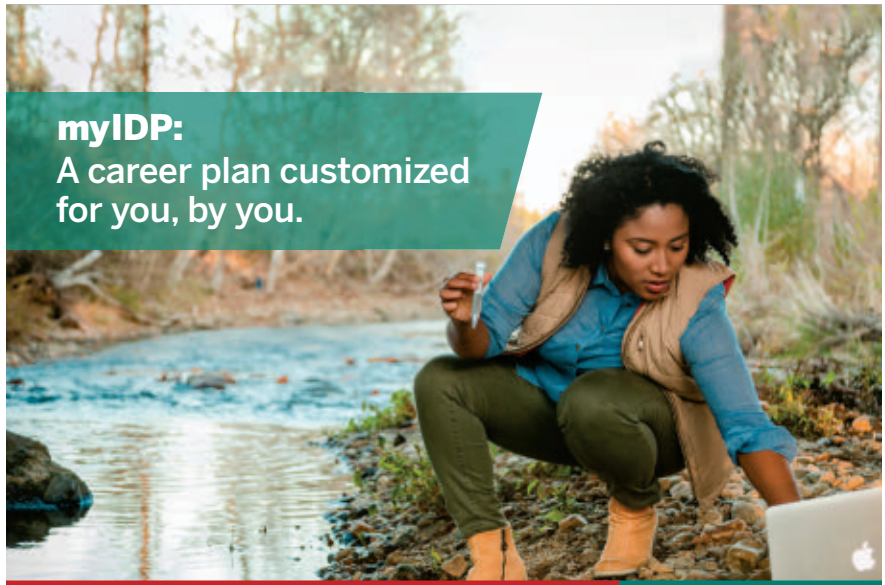
Established in 1898, RPCCC is the world's oldest cancer center and is known for numerous basic science and translational discoveries in oncology. RPCCC is an NCI-recognized Comprehensive Cancer Center that runs multidisciplinary research, combining, under the umbrella of its Cancer Center Support Grant, multiple Programs aimed at understanding, preventing and curing cancer. It has state-of-the-art research infrastructure enabling all aspects of modern biomedical research – from discovery work towards clinical validation – and provides a highly stimulating academic and clinical environment.

RPCCC is an equal opportunity employer that is committed to diversity and inclusion in the workplace. We prohibit discrimination and harassment of any kind based on race, color, sex, religion, sexual orientation, national origin, disability, genetic information, pregnancy or any other protected characteristic as outlined by federal, state or local laws. RPCCC makes hiring decisions based solely on professional qualifications, merit and specific research needs at the time.

Interested candidates should apply at: https://bit.ly/CellStressBiology

ROSWELL PARK COMPREHENSIVE CANCER CENTER

RPCCC and the University of Buffalo are M/F/D/V, M/F/H/V, Equal Opportunity/Affirmative Action Employer



For your career in science, there's only one Science

Features in myIDP include:

- Exercises to help you examine your skills, interests, and values.
- A list of 20 scientific career paths with a prediction of which ones best fit your skills and interests.
- A tool for setting strategic goals for the coming year, with optional reminders to keep you on track.
- Articles and resources to guide you through the process.
- Options to save materials online and print them for further review and discussion.
- Ability to select which portion of your IDP you wish to share with advisors, mentors, or others.
- A certificate of completion for users that finish myIDP.

myIDP

Visit the website and start planning today!
myIDP.sciencecareers.org

ScienceCareers In partnership with:

UCSF University of California San Francisco

UIC

BURROUGHS WELLCOME FUND

FASEB Federation of American Societies for Experimental Biology

University of Massachusetts Medical School UMASS

Director, Lunenfeld-Tanenbaum Research Institute
Vice President, Research, and Lou Siminovitch Research Chair
Sinai Health System, Toronto, Canada

Applications are invited for the Director, Lunenfeld-Tanenbaum Research Institute; Vice President, Research; and inaugural Lou Siminovitch Research Chair at Sinai Health System. Sinai Health is a fully-affiliated academic hospital of the University of Toronto, one of the world's top ranked research universities. We are seeking an experienced leader, able to develop and implement a strategic research vision, build and renew our research enterprise, and guide the research priorities for our institution.

The Lunenfeld-Tanenbaum Research Institute (LTRI) at Sinai Health is a centerpiece of Canadian biomedical enterprise, integrating a strong discovery research hub with impactful clinical programs. In addition to the LTRI, Sinai Health includes Mount Sinai Hospital, Hennick Bridgepoint Hospital and a system partner, Circle of Care. The LTRI is on the brink of an exciting future with planned expansion to a new physical space, recruitment of additional investigators at different career stages, and the opportunity to rethink and rebrand its identity under new leadership.

The LTRI is led by the Director, who reports to the President & CEO. It is comprised of 130 scientists and clinician-scientists, 450 research technical/administrative staff and 250 research trainees. Individual researchers have international reputations for excellence, as demonstrated by awards, large scale funding, and publications in the major areas of system biology, stem cell, genetics & heredity, precision medicine and quantitative data science, with applications to clinical oncology, obstetrics & gynecology, endocrinology & metabolism and cell biology. The LTRI is well equipped with major investments in robotics, imaging, animal research, biorepositories and tissue banks.

The successful candidate must hold a PhD, MD/PhD or MD degree and be eligible for fulltime academic appointment at the rank of Associate or Full Professor at the University of Toronto. The preferred candidate will have an international reputation for excellence in biomedical research. They will have the ability to set the direction for research at Sinai Health, with experience and knowledge to develop a strategic vision that includes basic science, as well as translational, population health and clinical research. As Vice President, Research, they will have experience in leading in a complex environment, be committed to collaboration with a broad spectrum of stakeholders, and play a key senior leadership role at Sinai Health. They will have excellent interpersonal skills and demonstrated ability to create and maintain a workplace and learning environment that is equitable, respectful and safe for all staff and trainees.

Remuneration for the role will be competitive and commensurate with qualifications and experience, in the range of $400,000-$450,000 per annum. The position is for a 5-year term, renewable once, based upon a successful review. The effective start date is November 1, 2022 or thereafter.

Applications for this position will be accepted until July 31, 2022. Interested candidates should forward a cover letter, curriculum vitae and names of three referees to:

Dr. Gary Newton
President & CEO, Sinai Health System
600 University Avenue, Ste. 19-310
Toronto, ON M5G 1X5, Canada
Gary.Newton@sinaihealth.ca

Sinai Health is committed to fostering a culture and environment that is safe, equitable, compassionate and accountable. All qualified candidates are encouraged to apply; however, Canadians and permanent residents will be given priority.

ScienceCareers online @sciencecareers.org

Who's the top employer for 2021?



Science 2021 TOP EMPLOYER

Science Careers' annual survey reveals the top companies in biotech & pharma voted on by *Science* readers.

Read the article and employer profiles at sciencecareers.org/topemployers

國立臺灣大學 National Taiwan University

National Taiwan University Presidential Search

National Taiwan University (NTU) is a full-fledged comprehensive university with outstanding achievements. Our Presidential Search Committee now invites nominations and applications for the position of President.

Presidential candidates should meet the following requirements: Academician at Academia Sinica, or professor, or professional with prior teaching and academic research experience equivalent to that required of a professor, and at least three years of experience, accumulative, as a director in educational institutions, government agencies, or in other state-owned or private business entities; or have held or currently hold the position of president at a university prior to the enforcement of the amended Act Governing the Appointment of Educators, dated November 15, 2011; or meets the qualifications for appointment as president of a university prior to the amendment. The ideal President should be a leader of integrity with excellent academic achievement and administrative skills in education. The new leader should also be able to administer matters impartially and beyond the interests of any political groups.

For those who would like to nominate a candidate, please fill in all required forms and deliver them to the address below via registered mail, express mail, or in person **by 5:00 p.m. on Monday, June 27, 2022**:

NTU Presidential Search Committee
No.1, Sec. 4, Roosevelt Road
Taipei
Taiwan 10617

Please note that the nomination/application must reach the Committee by the above deadline. The application dossier will not be returned.

Direct any inquiries to:
Pei-Yu Sun and Yi-Chen Hu, Office of the Secretariat
Tel: +886(0)2-3366-2032, 3366-1483
Fax: +886(0)2-2362-9997
Email: ntupsc@ntu.edu.tw

The nomination forms can be downloaded from the NTU Website:
http://www.ntu.edu.tw

National Taiwan University Presidential Search Committee

Shenzhen Institute of Advanced Technology
Chinese Academy of Sciences



Established in partnership between the Chinese Academy of Sciences and the Shenzhen Municipal Government, the Shenzhen Institute of Advanced Technology (SIAT) is a newly-created university with an objective to become the world's preeminent institute for emerging science and engineering programs. SIAT is equipped with state-of-art teaching and research facilities and is dedicated to cultivating international, visionary, and interdisciplinary talents while delivering research support to pursue innovation-driven development.

SIAT is located in Shenzhen, also known as the "Silicon Valley of China," a modern, clean, and green city, well-known for its stunning architecture, vibrant economy, and its status as a leading global technology hub. SIAT is seeking applications for faculty positions of all ranks in the following academic programs: Computer Science and Engineering, Bioinformatics, Robotics, Life Sciences, Material Science and Engineering, Biomedical Engineering, Pharmaceutical Sciences, Synthetic Biology, Neurosciences, etc. SIAT seeks individuals with a strong record of scholarship who possess the ability to develop and lead high-quality teaching and research programs. SIAT offers a comprehensive benefits package and is committed to faculty success throughout the academic career trajectory, providing support for ambitious and world-class research projects and innovative, interactive teaching methods.

Further information:
https://www.siat-sz.edu.cn/

What's Your Next Career Move?

From networking to mentoring to evaluating your skills, find answers to your career questions on *Science* Careers

To view the complete collection, visit
ScienceCareers.org/booklets



ScienceCareers
FROM THE JOURNAL SCIENCE AAAS

THE UNIVERSITY OF ARIZONA COLLEGE OF MEDICINE TUCSON

Associate or Full Professor, Immunobiology
College of Medicine, University of Arizona Health Sciences

The University of Arizona College of Medicine seeks applicants for a Tenure Track Associate Professor or Tenure Track Professor rank in Systems Biology. The successful candidate will hold an appointment in the Department of Immunobiology at the University of Arizona. This position will also be responsible for oversight of the UArizona Human Immune & Molecular Profiling facility. Within the department, the candidate will be expected to participate in medical and/or graduate teaching and service at the departmental, college and university levels. Teaching loads vary but are not expected to exceed 30 contact hours per year for research-intense faculty.

Applicants must have a PhD and/or MD, an established program in systems biology of the immune system, an active funded lab in immunology of infectious diseases or related area(s), and peer-reviewed publications in nationally recognized journals. Expertise and publications in single-cell omics will be a plus.

The successful candidate will interact with, and contribute 10% of time to, the strategic UAHS initiatives involving molecular profiling, including the initiative in Personalized Defense. That would include, but not be limited to, attending initiative meetings, and providing advice and expertise towards analysis of systems biology datasets. The candidate is expected to maintain a vibrant, extramurally funded research program with potential for translational applications; contribute to graduate (Ph.D.) and/or medical (M.D.) education; and provide institutional service in the Department of Immunobiology. Salary and start-up funds are attractive and commensurate with qualifications and experience.

Outstanding UA benefits include health, dental, and vision insurance plans; life insurance and disability programs; sick leave and holidays; UA/ASU/NAU tuition reduction for the employee and qualified family members; state and optional retirement plans; access to UA recreation and cultural activities; and more!

For full details and qualifications, and to complete an online application, see req9131 at talent.arizona.edu. The University of Arizona is an Equal Opportunity Employer Minorities/Women/Vets/Disabled.

ScienceCareers online @sciencecareers.org

By **Edgar Virguez**

What I promised my mother

"Daddy, you're on TV!" my daughter exclaimed, her face glowing with excitement and pride. We were at Duke University's Perkins Library, where I was featured in an exhibition about the institution's Latin community. It was a far cry from my first visit to the library 6 years earlier, soon after I came to the United States to pursue my graduate education. Walking into the majestic library for an orientation session back then, I quickly felt out of place when we were offered racial and ethnic categories to assign ourselves to. I consider myself mestizo—of mixed Indigenous and European descent, like many Latin Americans. But mestizo wasn't an option, and none of the categories offered seemed to fit me. I felt profoundly out of place.

I grew up in a low-income family in Colombia, and my mother sacrificed everything to give my sister and me a chance to succeed. She reminded us that because of our station in life we were unlikely to get second chances, so we needed to make the most of every opportunity—and enable paths for others. "Remember your origins," she told us.

Those words gained new meaning when I arrived in the United States in 2016. My wife and I faced racism and xenophobia. As an international student, I was ineligible for many fellowships. Amid the complex and rapidly changing immigration system, I felt I was constantly looking over my shoulder, weighing whether each professional opportunity that came my way would affect my eligibility to stay in the country.

In the face of these challenges, I felt more strongly than ever that I should help open doors for others. I applied my engineering training, defining skills to acquire and milestones to meet. Like any graduate student, I wanted to expand my technical knowledge and secure fellowships. I also wanted to increase campus diversity and engage with activist groups. When I checked my progress after my first year, though, I was disappointed to find that I had fallen short of the inclusion goals. I was failing to honor the commitment I had made to my mother, and to myself.

Not long after, my first child was born, another reminder of my priorities. After finishing my paternity leave, I returned to work recommitted to fighting for a diverse, inclusive environment.

With my adviser's support, I set aside time for advocacy work every week. In the years that followed, I orchestrated an agreement between my school and a Colombian nonprofit to fund Colombian graduate students at Duke. I launched an undergraduate Spanish course that included a project connecting Duke students with children in marginalized communities in Bogotá. I participated in committees advocating for inclusive education.



"I felt more strongly than ever that I should help open doors for others."

It was not always easy—or possible—to make headway on my advocacy as well as my research, and at times I didn't get the balance quite right. But when I veered too far in either direction, my adviser, an extraordinary mentor who listened and supported my goals, helped me find my way back. And when one of my thesis committee members suggested I decrease my engagement in nonresearch activities, my adviser listened to my frustration about how some people don't understand what it means to be oppressed.

I also found that some of the skills I developed fed back into my research. For example, I discovered that by dissecting my university's current practices, I could identify ways the institution could better support international students. That same approach helped me formulate the analysis for one of my research questions about where to build solar farms.

Just before defending my dissertation in January, I visited the library exhibition for the first time. I had been interviewed about the course I created, and I was overwhelmed to find myself featured. And when I looked up, a proverb emblazoned over the exhibition brought me to tears: "They tried to bury us; they didn't know we were seeds." ■

Edgar Virguez is a postdoctoral scholar at Carnegie's Department of Global Ecology at Stanford University. Send your career story to SciCareerEditor@aaas.org.

ILLUSTRATION: ROBERT NEUBECKER

science.org/journal/sciimmunol



READY TO PUT THE SPOTLIGHT ON YOUR RESEARCH?

Submit your research:
cts.ScienceMag.org

Science Immunology
AAAS

Twitter: @SciImmunology
Facebook: @ScienceImmunology

NOMIS & Science
YOUNG EXPLORER AWARD 2022

Research at the intersection of the social and life sciences

Unconventional. Interdisciplinary. Bold.

The **NOMIS & *Science* Young Explorer Award** recognizes and rewards early-career M.D., Ph.D., or M.D./Ph.D. scientists that perform research at the intersection of the social and life sciences. Essays written by these bold researchers on their recent work are judged for clarity, scientific quality, creativity, and demonstration of cross-disciplinary approaches to address fundamental questions.

A cash prize of up to 15 000 USD will be awarded to essay winners, and their engaging essays will be published in ***Science***. Winners will also be invited to share their work and forward-looking perspective with leading scientists in their respective fields at an award ceremony as well as a meeting of the NOMIS Board of Directors to consider future funding.

Apply by May 15, 2022
at www.science.org/nomis

NOMIS FOUNDATION
Creating the Spark

Science
AAAS